THE UNIVERSITY
of EDINBURGH

Genomics of divergence and hybridisation in the genus *Antirrhinum* L.

Mario Saturnino Duran Castillo

THE UNIVERSITY *of* EDINBURGH

Dissertation submitted for the degree of Doctor of Philosophy

Institute of Evolutionary Biology

The University of Edinburgh

February 2019

Supervisors:

Dr. Alex Twyford

Dr. Andrew Hudson

Declaration

I hereby declare that this dissertation is an original piece of work that embodies the results of my own research. All work contained herein has not been submitted for any other degree. I acknowledge the nature and extent of work carried out in collaboration with others below:

Chapter 2: Samples were collected and processed by Andrew Hudson and Yvette Wilson in 2007 and 2009. Species with endemic or limited ranges were provided by Professor I. Mateu-Andres (Universidad de Valencia, Spain). Samples from the species *A. molle*, and *A. braun-blanquetii* were collected and donated by T. Gübitz. Accessions from the taxa *A. pseudomajus*, *A. striatum*, *A. siculum* and *A. latifolium* were provided by Professor E. Coen (JIC, Norwich) and C. Thébaud (CNRS, University of Paul Sabatier, Toulouse, France). Alignment data from the species *M. orontium* was provided by Annabel Whibley (JIC, Norwich). DNA extractions and library preparations were done by Yvette Wilson and Andrew Hudson.

Chapter 3

Samples were collected and processed by Andrew Hudson and Yvette Wilson in 2007 and 2009. DNA extractions and library preparations were done by Yvette Wilson and Andrew Hudson.

Chapter 4

Samples were collected by Andrew Hudson and Yvette Wilson in 2009. They are also responsible for the design and construction of the F2 cross used for the QTL analysis. Yvette Wilson is responsible for the data collection, specifically of the measurements of all the F2 individuals. Matthew Barnbrook performed the initial stages of the analysis of genetic data including the filtering of individuals according to missing

data and the transformation of RAD markers into multilocus markers. Unpublished material from Feng *et al.*, (2009) was provided by Andrew Hudson.

Signature of candidate:

Mario Saturnino Durán Castillo

Acknowledgements

I would like to thank my sponsor CONACYT for paying my living costs and tuition fees during this last four years. I would also like to express my deep gratitude to the Mexican people. First, because the money from their taxes allowed to study what I love. Second because they taught me the intricate relationship between plants and people. Because it was by talking with people from small communities that my journey as a botanist started.

I would like to thank my supervisor Andrew Hudson and all the members of his lab that helped extracting DNA, and preparing the RAD sequencing libraries for the data generated in chapters 2, 3, and 4. I would also like to thank Andrew for giving the chance of doing this project, for always giving me good and honest advice and for always having time to meet and share ideas with me.

Special thanks to Yvette Wilson and Matthew Barnbrook, who were involved in the collection and analysis of the data for chapters 2, 3 and 4.

I am particularly grateful with my main supervisor Alex Twyford for helping me in all the stages of the PhD. Alex thank you very much for accepting me as your student, for always having a positive attitude, for the patience shown, for all the comments, the advice and motivation during the last four years.

I would also like to thank all the amazing people I have met in the two institutes I had the pleasure to work. First, a big thanks to the people from the "big lab" that made my first year of PhD such an enjoyable experience. Then, all my friends in the Institute of Evolutionary Biology. Thank you for keeping me sane.

I also want to thank my family for all the unconditional love and support from the beginning of this crazy adventure. Once again thank you for teaching me to love nature and science.

Finally thank you Zapata. You have been my strongest support during these years. Thank you for being the greatest positive influence in my life and for all the infinite talks about plants. Thanks for sharing this experience with me. Thank you for being here*. Porque en la calle codo a codo somos mucho más que dos.*

Oda a la crítica

Yo escribí cinco versos:
uno verde,
otro era un pan redondo,
el tercero una casa levantándose,
el cuarto era un anillo,
el quinto verso era
corto como un relámpago
y al escribirlo
me dejó en la razón su quemadura.

Y bien, los hombres,
las mujeres,
vinieron y tomaron
la sencilla materia,
brizna, viento, fulgor, barro, madera
y con tan poca cosa
construyeron paredes, pisos, sueños.
En una línea de mi poesía
secaron ropa al viento.
Comieron
mis palabras,
las guardaron
junto a la cabecera,
vivieron con un verso,
con la luz que salió de mi costado.
Entonces
llegó un crítico mudo
y otro lleno de lenguas,
y otros, otros llegaron
ciegos o llenos de ojos,
elegantes algunos
como claveles con zapatos rojos,
otros estrictamente
vestidos de cadáveres,
algunos partidarios
del rey y su elevada monarquía,
otros se habían
enredado en la frente
de Marx y pataleaban en su barba,
otros eran ingleses,
y entre todos
se lanzaron
con dientes y cuchillos,
con diccionarios y otras armas negras,
con citas respetables,
se lanzaron
a disputar mi pobre poesía
a las sencillas gentes
que la amaban:
y la hicieron embudos,
la enrollaron,
la sujetaron con cien alfileres,
la cubrieron con polvo de esqueleto,
la llenaron de tinta,
la escupieron con suave
benignidad de gatos,
la destinaron a envolver relojes,

la protegieron y la condenaron,
le arrimaron petróleo,
le dedicaron húmedos tratados,
la cocieron con leche,
le agregaron pequeñas piedrecitas,
fueron borrándole vocales,
fueron matándole
sílabas y suspiros,
la arrugaron e hicieron
un pequeño paquete
que destinaron cuidadosamente
a sus desvanes, a sus cementerios,
luego
se retiraron uno a uno
enfurecidos hasta la locura
porque no fue bastante
popular para ellos
o impregnados de dulce menosprecio
por mi ordinaria falta de tinieblas,
se retiraron todos y entonces,
otra vez,
junto a mi poesía
volvieron a vivir
mujeres y hombres,
de nuevo hicieron fuego,
construyeron casas,
comieron pan,
se repartieron la luz
y en el amor unieron
relámpago y anillo.
Y ahora,
perdonadme, señores,
que interrumpa este cuento
que les estoy contando
y me vaya a vivir
para siempre
con la gente sencilla.

Pablo Neruda

Lay summary

One of the main goals in evolutionary biology is to understand the mechanisms driving the origin of new species. Traditionally, the speciation process is considered to be the result of an interplay between forces originating new diversity (mutation and natural selection) and forces homogenising populations (gene flow, recombination). In recent decades the development of new sequencing technologies and the study of genomes from a wide variety of organisms have resulted in the development of new paradigms about the origin of divergence between populations in the presence of gene flow. However substantial challenges remain in elucidating the role of different evolutionary processes in the generation of new species. The genus *Antirrhinum*, including the model system *A. majus*, has many resources for addressing a wide range of evolutionary questions about speciation. Here I identify patterns of hybridisation and divergence in the genome of *Antirrhinum* species using a next generation sequencing approach. First, I reconstruct the phylogeny and explore the genetic structure and patterns of morphological evolution of the genus. I find evidence of hybridisation and the independent evolution of similar morphologies across the genus. Second, I investigate the extent of hybridisation in putative natural hybrid populations between species from different subsections (*Antirrhinum* and *Kickxiella*) and characterised the genetic structure of populations in the South of Spain. I found a high proportion of hybrid individuals in the populations in the south of the Sierra Nevada and identified the role of this mountain range as an important barrier to gene flow. Third, I explored the genetic architecture of traits that differ between subsection *Antirrhinum* and *Kickxiella* and identified important genomic regions for the maintenance of species. In this thesis I provide multiple lines of evidence suggesting that natural selection via habitat preferences plays a major role in maintaining the morphological divergence of species in the wild.

Abstract

Plant's sessile habit and extreme variation in ploidy level, mating system and dispersal provide excellent opportunities for the study of speciation. However substantial challenges remain in elucidating the role of different evolutionary processes in the generation of new species. The genus *Antirrhinum*, including the model system *A. majus*, has many resources for addressing a wide range of evolutionary questions about speciation. Here I use restriction site associated DNA (RAD-seq) to investigate the genomics of hybridisation and divergence in the genus *Antirrhinum*. First, I examined the phylogenetic relationships within the genus and estimate genetic structure and rates of hybridisation. My results show that geography has a strong influence on the phylogenetic relationships of the genus, and that hybridisation occurs repeatedly between the main taxonomic subsections. However I also found differences between the genetic clustering and the traditional taxonomy of the genus, with subsection *Kickxiella* appearing in several places in the tree. I recovered a *Kickxiella* like morphology as the ancestral state of the genus, and found several transition events to an *Antirrhinum* like morphology throughout the genus. Second, I measured the proportion of hybridisation in putative natural hybrid populations between species from different subsections (*Antirrhinum* and *Kickxiella*) and characterised the genetic structure of several populations in the South of Spain. I found a high proportion of hybrid individuals in the populations in the south of the Sierra Nevada and identified the role of this mountain range as an important barrier to gene flow. Third, I explored the genetic architecture of the morphological divergence between subsection *Antirrhinum* and *Kickxiella* and identified important genomic regions for the maintenance of species. In this thesis I provide multiple lines of evidence suggesting that natural selection via habitat preferences plays a major role in maintaining the morphological divergence of species in the wild.

List of contents

List of figures

List of tables

Chapter 1
General introduction

General introduction

Speciation

One of the main goals in evolutionary biology is to understand the mechanisms behind the origin and maintenance of biological diversity. Speciation, or the process of the origin of new species, is driven by two types of evolutionary forces (Slatkin, 1987; Coyne and Orr, 2004; Feder *et al*., 2012). On one hand, mutation, natural selection and genetic drift underlie the generation of new diversity and cause population differentiation (Lenormand, 2002). On the other hand, gene flow and recombination, as homogenising forces, constrain the evolution of organisms (Slatkin, 1987). Traditionally, since the introduction of the biological species concept by Mayr (1942), the study of speciation has focused on reproductively isolated species as the final product of the speciation process, and emphasized the importance of reproductive isolating barriers, with the aim of understanding their origin and underlying mechanisms (Bohonak, 1999; Coyne and Orr, 2004).

A more recent approach to the study of speciation is the study of species with incomplete reproductive barriers or hybridising populations to explore the evolution of reproductive isolation or the maintenance of species barriers (Noor and Feder, 2006). This research is driven by the emergence of new molecular techniques that allow large amounts of sequencing data to be generated for non-model species including those characterised by shallow species divergence where few nucleotide substitutions are fixed in a given lineage. These recent studies are exploring questions about the extent of divergence that can occur with gene flow, and the types of genes among taxa that can help or hinder speciation (Rieseberg and Blackman, 2010; Feder *et al*., 2012). From this research is now clear that in different organisms the interaction between ecology and genomic architecture can have a major role in speciation, and that gene flow is common in all stages of the speciation process (Huber *et al*., 2015; Marques *et al*., 2016; Price *et al*., 2018).

Evolutionary radiations

An evolutionary radiation is a rapid series of speciation events that result in an exceptionally diverse clade (example figure 1.1). Hence studying species radiations is fundamental to understanding patterns of biological diversity and their mode of evolution (Neige, 2015). Although there is a major discussion in the literature about what aspects make a true radiation event (Givnish, 2015), most opinions agree that radiations have a common ancestry (monophyly), an increased diversification rate and an adaptive component (Hughes *et al*., 2015). This definition of radiation assumes that natural selection will drive divergence between different populations capable of exploiting different ecological niches. Despite the existence of many examples of adaptive radiations in different organisms [e.g. Hawaiian silverswords (Baldwin and Sanderson, 1998); sticklebacks (Drummond *et al*., 2012), Darwin finches (Wagner *et al*., 2013), and cichlid fish (Brawand *et al*., 2015)], the mechanisms behind the diversification in adaptive radiations remain poorly understood (Drummond *et al*., 2012).

A large amount of our current knowledge about adaptive radiations comes from macro-evolutionary studies. These studies, based on phylogenetics and comparative biology, have contributed to our understanding of spatial patterns and ecological factors in species diversification over long time scales (Glor, 2010; de la Harpe *et al*., 2017). A significant advance in this field has been the development of molecular time calibrated phylogenies that allow the comparison of diversification rates and the identification of rate shifts across phylogenies (Hasegawa *et al*., 1985; Richardson, Pennington, *et al*., 2001; Hughes *et al*., 2015). Another important advance has been the application of techniques based on the principles of population genetics for the dissection of the processes of local adaptation and speciation (Hodges and Derieg, 2009; de la Harpe *et al*., 2017). In particular, methods based on the allele frequency spectrum, coalescent theory, and the estimation of admixture have increased our understanding of the prevalence of gene flow in all stages of the speciation process while also resolving the demographic speciation histories of a wide variety of organisms (Campbell *et al*., 2018). For example, studies of animals like African

cichlid fish (Brawand *et al*., 2015), Darwin's finches (Petren *et al*., 2005) and *Heliconius* butteflies (Nadeau *et al*., 2013), and plants like oak trees (Cavender-Bares *et al*., 2015) and white proteas (Prunier and Holsinger, 2010) have not only identified important geographic and ecological factors occurring in radiations, but also identified major evolutionary forces shaping the genetic structure within and between populations. Overall these studies have showed the importance of adaptive and non-adaptive forces in speciation.

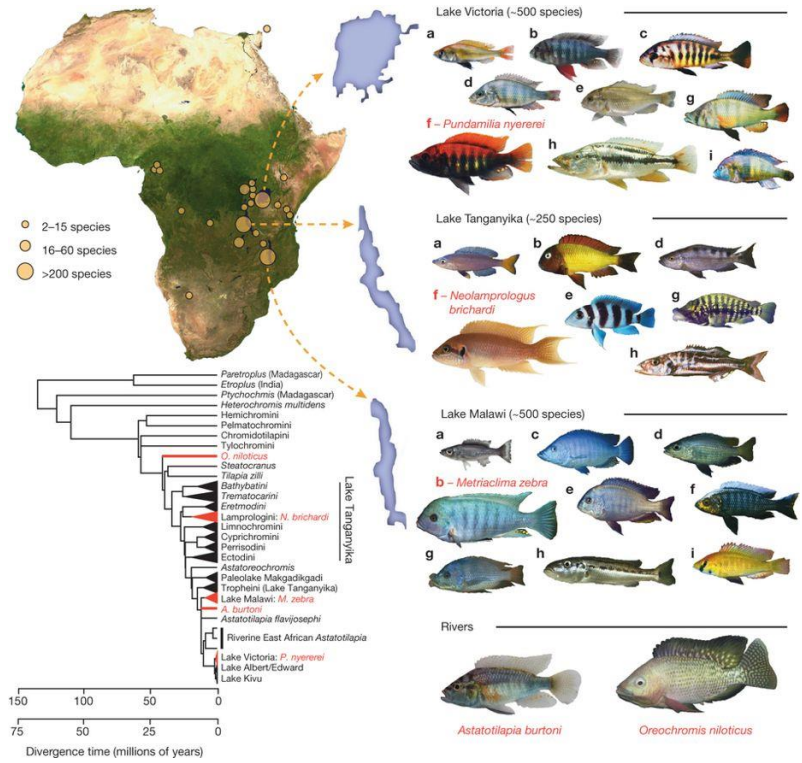


Figure 1.1. The adaptive radiation of African cichlid fish. On the top left, map showing the lakes where cichlid fish have radiated. On the right, examples of morphological variation in cichlids. The names in red correspond to the species used on the study. On the bottom left a phylogeny of the major lineages and major radiations [from Brawand *et al*. (2015)].

Hybridisation and hybrid zones

Hybridisation is widely recognized as an important phenomenon in the evolution of plants and animals (Mallet, 2007; Soltis and Soltis, 2009; Abbott *et al*., 2013). In some cases, hybridisation can be an important source of new variation that can lead to speciation by two mechanisms. The first one is allopolyploid speciation (Coyne and Orr, 2004) and involves hybridisation and genome duplication with the hybrid offspring developing reproductive isolation almost immediately. This type of speciation is well characterised in plants like *Nicotiana tabacum*, *Iris versicolor* and *Gossypium hirsutum* (Stebbins, 1950). The second scenario has been called homoploid hybrid speciation (Rieseberg, 1997; Mallet, 2007) and involves species with equal chromosome number originating progeny with the same chromosomal number by hybridisation. In some circumstances, the origin of chromosomal rearrangements after hybridization can lead to reproductive isolation between the hybrid offspring and the parental species in a process called recombinational speciation (Coyne and Orr, 2004). Probably the best documented case of homoploid speciation are the sunflower species *Helianthus anomalus*, *H. paradoxus* and *H. deserticola*, in which several chromosomal rearrangements and the colonization of new ecological niches cause reproductive isolation from the parental species *H. annuus* and *H. petiolaris* (Rieseberg *et al*., 1995).

During a hybridisation event, the genetic information of each parental form interacts to create novel genotypic combinations in the offspring. This interactions can have different effects on the phenotypes of the new hybrids that go from producing offspring with intermediate morphologies to producing individuals with extreme phenotypes (transgressive segregation). Depending on the phenotypes produced and the interaction with the environment, certain genotypes will be removed from the hybrid population by natural selection (Abbott *et al*., 2013). At the genetic level, this selection of hybrid genotypes will cause differential introgression in the progeny. This means that some alleles at some loci will introgress more frequently than others. For example, alleles responsible for local adaptation will be selected

against in the hybrid progeny and will not introgress while neutral alleles will introgress at different degrees (Harrison and Larson, 2014). This process leaves a distinctive pattern in the genome that can be detected analysing the genomic patterns of variation between individuals and populations (Malek *et al*., 2012; Nosil *et al*., 2012; Lindtke *et al*., 2013; Renaut *et al*., 2014).

Natural hybrid zones have been called natural laboratories (Hewitt, 1988) for studies of evolution and speciation and they offer several advantages for the research of hybridization. First, they are very common in nature with many reports of hybrid zones in animals and plants (Hewitt, 1988; Abbott, 2017); Second, population genetic studies of hybrids zones can help us to understand the interplay between selection, gene flow, ecology and demography in maintaining the genetic integrity of the species (Noor and Feder, 2006). Third, they allow for the detection of candidate "barrier genes", when regions with low levels of introgression and high levels of divergence are present despite hybridisation (Noor and Feder, 2006). Finally, hybrid zones contain a great number of recombinant individuals as result of several generations of natural interbreeding and therefore they are a remarkable resource for the genetic mapping of traits in non-model organisms (Rieseberg and Buerkle, 2002).

Genomics

In the last few decades, there has been a substantial progress in speciation research mainly due to the recent genomic revolution (Feder *et al*., 2012; Campbell *et al*., 2018). A reference genome or a closely related model species are no longer required for the generation of genomic data in non-model organisms. This allows us to use thousands or millions of genetic markers to study species in problematic taxa (e.g. in recent radiations or experiencing gene flow) and increasing our ability to study not only ecological and geographical processes but also differences in the genome structure of individuals and divergent populations (Campbell *et al*., 2018). Advances in genomics has also resulted in improvements in the genotyping of experimental

crosses, thus increasing the efficacy of genome-wide association studies and quantitative trait loci analyses (Noor and Feder, 2006).

Reduced representation libraries (RRL) are cost efficient approaches to sample a small percentage of the genome using new generation sequencing (NGS) technologies. These methods consist of the digestion of many DNA samples with restriction enzymes followed by the sequencing of the regions flanking restriction sites of all or a subset of the restriction fragments. Probably the most widely used RRL is restriction site associated DNA sequencing (RAD sequencing; figure 1.2). Unlike other methods to generate genomic data, RAD sequencing does not require previous genomic knowledge of the system and therefore has become a popular approach for the discovery and genotyping of thousands of markers in non-model organisms. RAD sequencing also provides great flexibility in the number of markers produced by selecting the right restriction enzyme and the right method from a wide variety of techniques created based on the original RAD protocol (Andrews *et al*, 2016). RAD markers have been used successfully in many organisms to answer different questions about ecology, phylogenetics and speciation (Emerson *et al*., 2010; Rubin *et al*., 2012; Jones *et al*., 2013; Nadeau *et al*., 2013; Wagner *et al*., 2013b). In this work I use restriction site associated DNA markers (RAD) to disentangle the relationships in a recent radiated clade, to explore the genetic structure of populations in a hybrid zone and finally, to explore the genomic architecture of adaptive traits that are divergent between species.

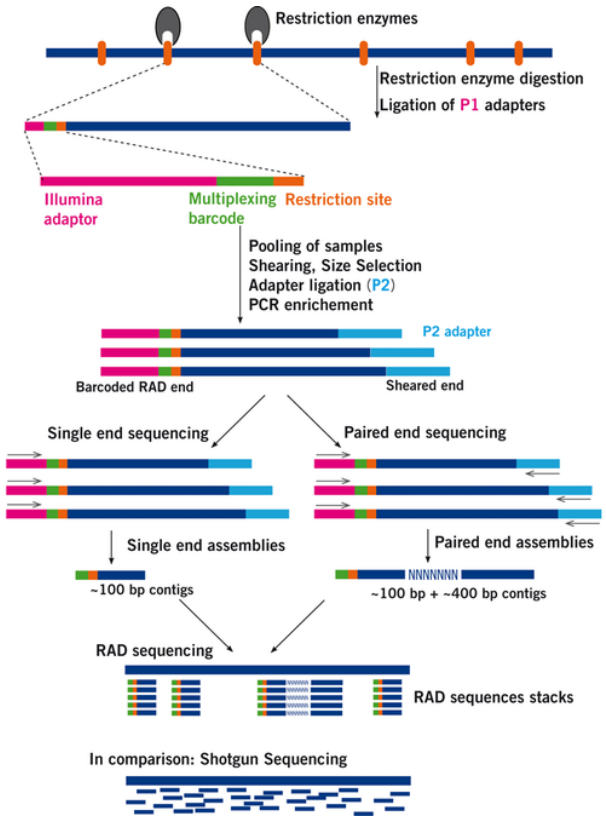


Figure 1.2 Diagram of the original RAD sequencing protocol. On top, DNA is digested with a restriction enzyme followed by the attachment of barcoded P1 adaptors. Then, samples are pooled and the DNA is sheared and filtered according to size. P2 adaptors are ligated to the sheared ends and only fragments with the two adaptors are amplified with a PCR. Finally, samples are sequenced using and NGS platform. The resulting reads can be single end, if fragments are sequenced only from one extreme, and paired end, if fragments are sequenced from both extremes.
Bottom, comparison with shotgun sequencing (taken from https://www.floragenex.com/rad-seq/).

Genomic architecture of speciation

During speciation, the patterns of divergence across the genome can be highly heterogeneous. For example, independent evolution of species in allopatry will promote the accumulation of divergence in different loci in each population. Other factors like genetic drift, gene flow, mutation rate and the distribution of genes under selection and their effect size can also impact genomic divergence during speciation

(Nosil *et al*., 2009). These patterns of heterogeneous genome divergence are known as the genomic architecture of speciation and the field of population genomics aims to dissect and identify their main causes and consequences on the evolution of organisms (Nosil and Buerkle, 2010).

Studying the genomic architecture of speciation has increased our understanding of the mechanisms of speciation in the absence of strong geographic isolation. Genomic Islands of speciation are genomic regions under strong divergent ecological selection that show patterns of divergent evolution even in the presence of gene flow (Turner *et al*., 2005; Campbell *et al*., 2018). These divergent regions, which underlie reproductive barriers or adaptations, stand out in comparison to the rest of the genome that should be homogenised by gene flow. Additionally, these effects are not limited to the loci under strong selection but also to areas in close proximity or linked in a process called genomic hitchhiking or divergent hitchhiking (Via, 2012; Ravinet *et al*., 2017). The size of the region affected by hitchhiking will depend on factors like the recombination rates and the strength of selection (Nosil *et al*., 2009). For example, stronger divergent selection and lower recombination rates will increase the effects of genetic hitchhiking producing bigger islands of divergence. While influential, the simple verbal model of islands of speciation remains contentious, and further theoretical and empirical results are required to explain how these islands grow, or indeed whether this model fits natural populations.

Relative measures of diversity, such as Fst, are usually used to document the degree of divergence and to identify these "islands". However, the interpretation of these genomic outliers has proven to be difficult and the assumption that loci with high Fst values are under divergent selection has to be taken with caution (Cruickshank and Hahn, 2014). Therefore, other summary statistics not dependant on levels of within population diversity, such as Dxy, as well as other methods to test divergent selection on genomic regions have to be implemented. For example, an effective way to identify regions underlying adaptive divergence is compare QTL for traits under divergent selection to sites with high Fst in population samples (Hawthorne and Via,

2001; Brennan *et al*., 2016). Although the interpretation of genomic outliers has proven to be complicated, undoubtedly this approach has been valuable for driving empirical progress specially in the search for loci contributing to reproductive isolation (Cruickshank and Hahn, 2014; Ravinet *et al*., 2017). The characterisation of the number and location of these sites is still a main goal of many studies in model and non-model organisms(Hawthorne and Via, 2001; Turner *et al*., 2005; Martin *et al*., 2013; Marques *et al*., 2016; Ragland *et al*., 2017).

The genus *Antirrhinum*

All chapters in this thesis (2-4) contain research on the genus *Antirrhinum* at different taxonomic levels (whole genus, species and populations). The relevant material is provided in each chapter, and here I broadly summarise our current knowledge of the diversity of *Antirrhinum*.

The genus *Antirrhinum* is a plant model system widely used for the research of the genetic mechanisms of plant development and in particular floral development (Hudson *et al*., 2007). The genus is composed of between 17 and 26 herbaceous species, depending on the taxonomic criteria used (Rothmaler, 1956; Sutton, 1988), and is mainly distributed in the Iberian peninsula and the Mediterranean region (figure 1.3). In this work I follow the nomenclature used by Rothmaler (1956), considering only species from the Old World as part of the genus *Antirrhinum* (*Antirrhinum sensu stricto*).

The popularity of the plant as a model system is partly due to the diverse flower colour, with colours going from white to magenta and yellow. Other traits that make this genus attractive as a model system are a diploid genome, the high level of phenotypic vegetative variation, combined with hardiness, a relatively short generation time, ease of experimental pollination and a lack of post pollination reproductive isolating barriers between almost all the species (Schwarz-Sommer *et al*., 2003). Among the several genetic mechanisms underlying development that are well characterised in *Antirrhinum* are flower development and differentiation of

floral organs, asymmetry, and the role of pollen in gametophytic self–incompatibility (Coen and Meyerowitz, 1991; Luo *et al*., 1996; Hudson *et al*., 2008a). Additionally, *Antirrhinum* is also emerging as a promising system for studying the role of natural selection and hybridisation on the evolution of divergence on natural populations (Whibley *et al*., 2006; Bradley *et al*., 2017; Tavares *et al*., 2018).

The amount of knowledge available about *Antirrhinum* development is in contrast with the lack of information about phylogenetic relationships. Previous attempts of phylogenetic reconstruction in *Antirrhinum* using nuclear and plastic loci report low sequence divergence and incongruence between loci (Adeyanju, 2003; Chaffe, 2003; Gubitz *et al*., 2003; Vargas *et al*., 2004, 2009; Jiménez *et al*., 2005; Wilson and Hudson, 2011). Despite low resolution and high homoplasy obtained with nuclear markers, haplotype distributions and network analyses show that geography has been a major factor structuring genetic diversity in *Antirrhinum* (Vargas *et al*., 2009; Wilson and Hudson, 2011). Dated molecular phylogenetic analyses of the genus suggest its evolution is characterised by a recent and rapid radiation with a crown age of divergence between 4.10 and 26 million years ago (Vargas *et al*., 2009, 2013). In addition, based on the distribution of important morphological characters in *Antirrhinum* and the limited distribution of several species, Rothmaler (1956) and Webb (1972) proposed multiple secondary contact events as a result of climatic fluctuations during the Pliocene-Pleistocene.

Traditionally the genus has been divided into three taxonomic subsections based on morphology: *Kickxiella*, *Antirrhinum* and *Streptosepalum* (Rothmaler, 1956). Subsection *Kickxiella* is characterised by plants with a prostrate habit, round leaves, stems full of trichomes and small, axillary, white flowers. Species within subsection *Antirrhinum* are taller with an upright habit, have longer and thinner leaves and lack glandular hairs outside the terminal inflorescence of magenta flowers. Plants in subsection *Streptosepalum* are tall with an upright habit, have long and thin leaves, with glandular hairs only in the inflorescence and are characterised by big yellow flowers arranged in a terminal inflorescence. Previous studies have linked these

differences in morphology with ecological factors. For example, subsection *Kickxiella* grows on cliffs and rocky surfaces with low vegetation cover, subsection *Antirrhinum* grows in deeper soils surrounded by a higher vegetation cover and subsection *Streptosepalum* is restricted to serpentine soils (García-Barriuso *et al.*, 2012; Wilson *et al.*, 2016).

There are now many resources available for research in *Antirrhinum*, including inbred laboratory lines from cultivars, a significant collection of mutants, cDNA libraries, a molecular linkage map (Schwarz-Sommer *et al*., 2003; Zhang *et al*., 2005) and a reference genome (Schwarz-sommer *et al*., 2003; Schwarz-Sommer *et al*., 2003; Hudson *et al*., 2007; LI *et al*., 2018). These genomic resources make the genus *Antirrhinum* ideal for tackling questions about the ecological and genomic processes underlying speciation in natural populations.

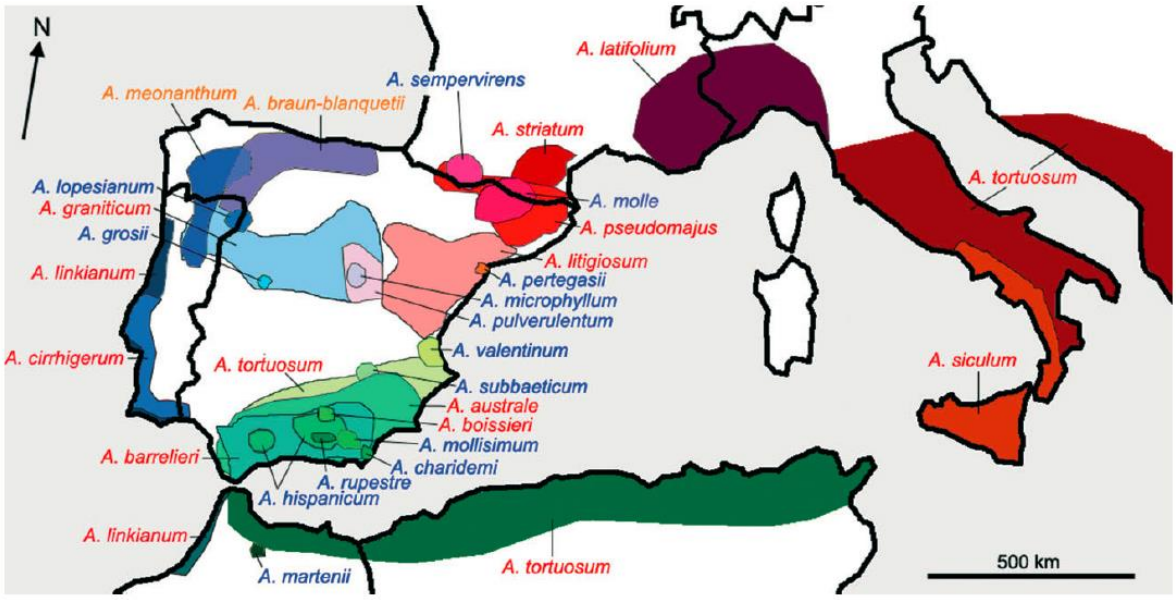


Figure 1.3. Distribution of the genus *Antirrhinum* in the Iberian Peninsula and the Mediterranean region (from Wilson & Hudson, 2011).

Aims and approaches

The main goal of this thesis is to investigate the genomics of divergence and hybridisation in the recently radiated genus *Antirrhinum*, with the aim of better understanding the speciation process in a rapid recent radiation. First, I explore the nature of species in *Antirrhinum* from a genomic perspective with the aim of producing a framework phylogeny for hypothesis testing (Chapter 2). Phylogenetic reconstructions of recent radiated taxa is known to be a challenging task caused by the lack of enough genetic and morphological differences between taxa. In addition, hybridisation and incomplete lineage sorting are common in these clades causing phylogenetic incongruence. Here I use a reduced representation sequencing approach (RAD sequencing) to resolve the phylogenetic relationships of *Antirrhinum* for the first time. Then I relate phenotypic variation to this new phylogenetic tree to reconstruct important transitions in morphology through the evolution of the genus. After resolving the phylogenetic relationships, I focus on a single clade that includes representatives from subsection *Antirrhinum* and *Kickxiella* that may have been characterised by hybridisation. These species are distributed in the South of Spain and co-occur in the Sierra Nevada and adjacent regions. Previous nuclear and plastid genotyping showed a strong genetic similarity between populations and the presence of haplotypes corresponding to the subsections *Antirrhinum* and *Kickxiella* (Wilson and Hudson, 2011). Hence these populations represent a good system to study the genome-wide patterns of introgression and divergence and the maintenance of species barriers in the presence of gene flow. I aim to identify these patterns of hybridization in natural hybrid populations using RAD markers (Chapter 3).

Finally, I aim to understand the genetic architecture of the morphological divergence between subsections *Antirrhinum* and *Kickxiella* (Chapter 4). Here I use a cross between the species *A. barrelieri*, with *Antirrhinum* morphology, and *A. rupestre*, with *Kickxiella* morphology, to perform a quantitative trait loci analysis (QTL) using RAD sequencing markers. I characterize the number of regions responsible for

morphological characters, their effect size and their location to assess the influence of the genomic architecture in the divergence of the two main subsections of *Antirrhinum*.

Chapter 2

The phylogeny of *Antirrhinum* reveals multiple transitions in flower colour and morphology

The phylogeny of *Antirrhinum* reveals multiple transitions in flower colour and morphology

Abstract

The study of recent and rapid radiations is important to understand the processes underlying the formation of biological diversity. A main issue in studying radiations is the low level of morphological and genetic diversity, and the occurrence of hybridisation and incomplete lineage sorting. Here I use restriction site associated DNA (RAD) markers to explore the nature of species and patterns of morphological evolution in the recently radiated plant genus *Antirrhinum*. First, I examined the phylogenetic relationships within the genus using maximum likelihood and coalescent methods. I found geography plays an important role in the divergence of the genus and that there are multiple occurrences of a *Kickxiella* morphology across different parts of the tree. Then, I estimated genetic structure and hybridisation rates at the whole genus level, and find evidence of hybridisation between members of different taxonomical subsections. Finally, I related morphological variation to the phylogenetic tree and performed an ancestral state character reconstruction. I recovered a *Kickxiella* morphology as ancestral for the genus and found several transition events to an *Antirrhinum* morphology throughout the phylogeny.

Introduction

Studying the processes underlying diversity in evolutionary radiations is fundamental to understanding past and current patterns of biological diversity. In some cases, radiations can be rapid and can be triggered by the origin of morphological novelty (e.g. wings) or ecological opportunity (e.g. the Andes uplift ; Richardson *et al.*, 2001; Drummond *et al.*, 2012; Givnish, 2015). These rapid radiation events have been suggested as one of the main mechanisms of organism diversifications (Schluter, 2009; Givnish, 2015). However, one of the main issues in the study of rapid radiations is to disentangle the relationship between species boundaries in taxa with low levels of genetic or morphological divergence. Additionally, hybridisation and incomplete lineage sorting are common processes in rapid radiations that add an extra level of complexity to the problem as the distribution of the genetic variation between taxa may follow the species tree. Only recently have new generation sequencing technologies allowed us to use thousands or millions of genetic markers to study the relationship between species. For example, RAD sequencing has been reported as promising for phylogenetic reconstructions of clades with a recent origin even in the absence of a reference genome (Rubin *et al*., 2012; Jones *et al*., 2013; Nadeau *et al*., 2013; Wagner *et al*., 2013b).

Flowering plants include many extraordinary examples of evolutionary radiations and much of this diversification has happened recently (Rieseberg and Willis, 2007). Perhaps the most remarkable outcome of radiations in flowering plants is the enormous variation of floral traits including variation in morphology, odour, colour and pigmentation patterns. In particular, flower colour has been found to evolve in response not only to pollinators but to biotic and abiotic interactions like herbivory and drought tolerance (Strauss and Whitall, 2006).

Flower colour is a complex trait with variation within species and populations in the pigments produced, their optical qualities and in the intensity and patterns of expression (for reviews on the topic see Rausher, 2008 and Glover, 2014). Despite

this large variation in the underlying expression patterns, plant pigments are produced by three main biochemical pathways: flavonoids, carotenoids and betalains. Historically a lot of attention has been given to flavonoids, as the pathway is highly conserved among land plants and they have a massive variation in functions from signalling to pollinators to UV protection as well as playing a fundamental role in the interaction between legumes and nitrogen fixing bacteria (Liu and Murray, 2016). Overall flower colour represent an excellent trait to test functional and morphological hypothesis of diversification during speciation events.

The genus *Antirrhinum* has been a plant model system for the study of mutation and inheritance since the last century because of their easy cultivation and the great variation in flower colour and morphology. Several genetic mechanisms common to angiosperms and involved in development have been well characterised in *Antirrhinum*. These include flower development and differentiation of floral organs, asymmetry and the role of pollen in gametophytic self–incompatibility (Coen and Meyerowitz, 1991; Luo *et al*., 1996; Hudson *et al*., 2008a). *Antirrhinum* has also played an important role in improving our understanding of the genetic mechanisms behind differences in flower colour and pigmentation patterns in the lab and natural populations (Bradley *et al*., 2017; Tavares *et al*., 2018).

The amount of knowledge available about *Antirrhinum* development is in contrast with the lack of information about the phylogenetic relationships. Previous attempts of phylogenetic reconstruction in *Antirrhinum* using nuclear and plastic loci report low sequence diversity and incongruence between loci (Adeyanju, 2003; Chaffe, 2003; Gubitz *et al*., 2003; Vargas *et al*., 2004, 2009; Jiménez *et al*., 2005; Wilson and Hudson, 2011). Despite a low resolution and high homoplasy obtained with nuclear markers, haplotype distributions and network analyses show that geography has been a major factor structuring genetic diversity in *Antirrhinum*. In fact, the limited distribution of many species made Rothmaler (1956), Webb (1971), Vargas *et al.* (2004) and Liberal *et al.* (2014) suggest that the divergence of the genus must have happened in the Iberian mountains by geographical isolation between 4.10 and 26

million years ago. This radiation was probably followed by multiple secondary contact events as a result of climatic fluctuations during the Pliocene-Pleistocene (Vargas *et al*., 2009).

Traditionally the genus has been divided into three taxonomic subsections based on morphology: *Kickxiella*, *Antirrhinum* and *Streptosepalum* (Rothmaler, 1956). Subsection *Kickxiella* is characterised by plants with a prostrated habit, round leaves, stems full of glandular trichomes and small, axillary, white flowers. Species within subsection *Antirrhinum* are taller with an upright habit, have longer and thinner leaves and lack hairs outside the terminal inflorescence of magenta flowers and first three internodes. Plants in subsection *Streptosepalum* are tall with an upright habit, have long and thin leaves, with glandular hairs only in the inflorescence and are characterised by large yellow flowers arranged in a terminal inflorescence (figure 2.1). Previous studies have linked these differences in morphology with ecological factors. For example, subsection *Kickxiella* grows on cliffs and rocky surfaces with low vegetation cover, subsection *Antirrhinum* grows in deeper soils surrounded by a higher vegetation cover and subsection *Streptosepalum* is restricted to serpentine soils in the north-west of the Iberian Peninsula (García-Barriuso *et al.*, 2012; Wilson *et al.*, 2016).

In this study, I test the nature of species differences in the recent radiated genus *Antirrhinum* from a genomic perspective. First I use a reduced representation sequencing approach (RAD sequencing) to resolve the phylogenetic relationships of closely related taxa and relate phenotypic variation to this new tree. Then I explore the distribution of traits and the lability of trait transitions across *Antirrhinum*. The recent origin of the genus and morphological specialisation to habitat type suggest there are ecological drivers of species diversification. I test the hypothesis that the different clades of *Antirrhinum* have evolved different suites of adaptive floral and growth-related traits that allow them to exploit new ecological opportunities, but that transitions of individual traits are maladaptive and rare in the wild. I predict that the ancestral phenotype of *Antirrhinum* had white flowers, and that new floral

phenotypes have evolved in each clade of *Antirrhinum*. I also estimate the extent of hybridisation to test whether this may spread adaptive genes between species. Here I provide the first insight into the phylogenetic relationships of *Antirrhinum* and propose an evolutionary theory for morphological evolution of important adaptive traits during a rapid radiation event.

Figure 2.1. *A. tortuosum* showing a typical *Antirrhinum* morphology (A, B). A. *menoanthum* showing the morphology of subsection *Streptosepalum* (C, D). *A. rupestre* show what is considered as a *Kickxiella* morphology (E, F). Picture A and B taken by Věra Svobodová and C and D were taken by Salvador Feo Garcia.

Methods

Specimen sampling and sequencing

Samples from 120 individuals from 28 *Antirrhinum* taxa were sampled by Andrew Hudson and Yvette Wilson in 2007 and 2009 across the entire range of the genus. Species with endemic or limited ranges were provided by Professor I. Mateu-Andres (Universidad de Valencia, Spain). Samples from the species *A. molle*, and *A. braun-blanquetii* were collected and donated by T. Gübitz. Accessions from the taxa *A. pseudomajus*, *A. striatum*, *A. siculum* and *A. latifolium* were provided by Professor E. Coen (JIC, Norwich) and C. Thébaud (CNRS, University of Paul Sabatier, Toulouse, France). The geographic coordinates for each accession are provided in table S.1.

DNA was extracted fresh tissue frozen at -80 °C or dried following a modified CTAB method. Around 1 mg of leaf tissue was ground for 2 minutes using a Retsch MM300 tissue lyser, and then with 600 µl of extraction buffer [100 mM Tris-HCL (pH 8.0), 1.4 M NaCl, mM EDTA, 2% CTAB and 0.2% MCE (mercaptoethanol)] for 1 minute. The ground sample was incubated at 65°C for 20 minutes, vortexed with 600 µl of chloroform and centrifuged at 12000 rpm for 10 minutes. The supernatant was transferred to a new tube adding 400 µl of isopropanol vortexed briefly and centrifuged at 18000 rpm for 10 minutes. The supernatant was pipetted off and the tube left open in a fume hood until all isopropanol evaporated. In order to remove RNA, 50 µl of TE buffer [10 mM Tris-HCl (pH 8.0), 1mM EDTA] with $1/1000^{th}$ RNAase was added to the tube and incubated at 4^0C overnight. DNA was then washed using $1/10^{th}$ the starting value of 3M NaOAc (pH5.2) and 2 ½ times the starting volume of ethanol and centrifuged at 20800 rpm for 10 minutes. The supernatant was removed and the tube was left open under a fume hood for around an hour until the ethanol evaporated completely. Finally, the DNA was dissolved in 50 µl of TE buffer.

Restriction-site associated DNA (RAD) libraries were prepared by Yvette Wilson by digesting each DNA sample with the enzyme *Pst* I. The fragments were then ligated to a sample specific P1 adapters with a 5'-TGCA-3' overhang ligated to it. These adaptors were bar-coded (e.g. the 5 nucleotides before the overhang differed). All

samples were pooled and sheared to around 250-550 bp and gel purified. Sheared ends were repaired with T4 DNA polymerase and a 3' overhanging A added with Klenow DNA polymerase. P2 adaptors were ligated to the sheared ends. These have a complementary overhanging T and were also bar-coded. In order to pool more sequences up to six different sets of P2 adaptors were used. The adaptor ligated fragments were then amplified with a few cycles of PCR using primers complementary to P1 and P2 adaptors. Finally, libraries were sequenced in an Illumina HiSeq-4000 at Edinburgh Genomics producing 104,031,701 paired end sequencing reads.

Processing and alignment of RAD-seq data

Raw reads were demultiplexed using the script *process_radtags* from the software Stacks (Catchen *et al*., 2013). The total number of reads sequenced per sample and the adaptors used are shown in table S.3. Adaptors were removed and reads trimmed and filtered by quality using the default settings in trimmomatic 0.36 (Bolger *et al*., 2014). Remaining reads were mapped to the version 2 *A. majus* genome using Bowtie2 (Langmead and Salzberg, 2012) and duplicated sequences removed using Picard tools(Broad Institute, 2018). Variant/invariant genotypes were called using samtools 1.6 and the multiallelic caller implemented in bcftools 1.4 (Li, 2011). This data set was then filtered by mapping quality, depth and missing data, where individuals with more than 70% missing data and loci with less than 3 of depth, 40 of mapping quality and more than 50% of missing data were removed. The final dataset was composed of 16,061,293 sites from 86 samples corresponding to 24 taxa (figure 2.2).

In order to root the phylogenomic trees I used whole genome sequence data from the species *Misopates orontium* that was kindly provided by Annabel Whibley from the John Innes Centre. Variant calling for this sample was done following the same pipeline used for the rest of the samples and the resulting vcf was filtered keeping only the loci present in the alignment of all *Antirrhinum* samples.

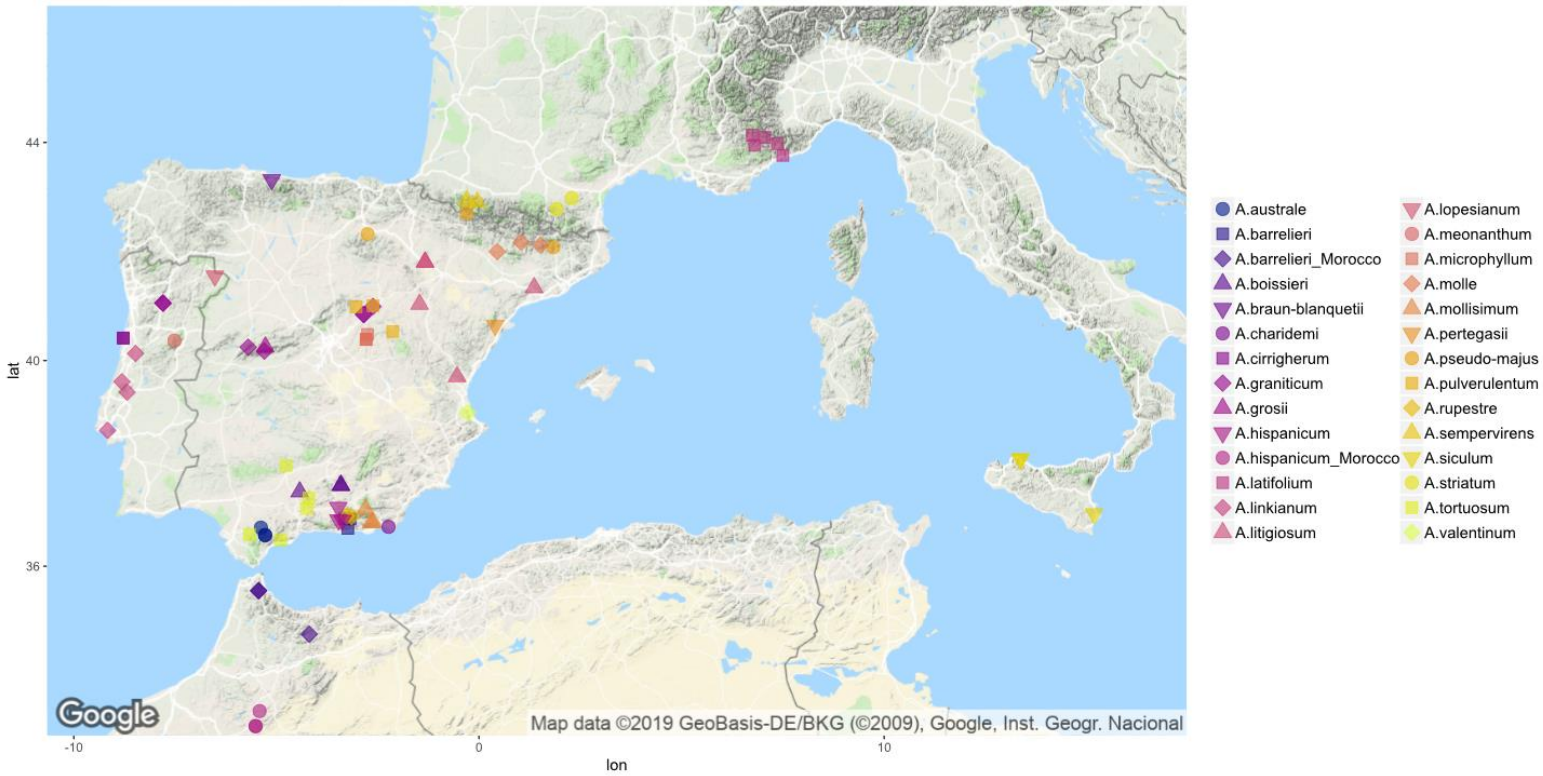


Figure 2.2. Geographic map showing the locations of the populations used for phylogenetic analyses.

Phylogenetic analyses

A maximum likelihood analysis was conducted on the concatenated dataset of variant and invariant sites using GTR-GAMMA substitution model implemented with RAxML (Stamatakis, 2014). Branch support was assessed using the rapid bootstrap analysis option with 100 replicates. In order to account for the presence of incomplete lineage sorting I used the software SVDquartets (Chifman and Kubatko, 2015) as an alternative tree estimation method under the multispecies coalescence. This software examines single site patterns in order to build unrooted species trees of four taxa at a time. Therefore the SVDquartets algorithm works by splitting the data into small groups of four taxa and assigning a score to each possible topology for each quartet. At the end of the analysis, the topologies with the lowest scores for each quartet are chosen and a conglomeration technique is used to combine each quartet tree into a species tree (Chifman and Kubatko, 2014).

The SVDquartets algorithm requires unlinked multi-locus data and therefore I filtered the previous alignment using the function *--thin* in vcftools (Danecek *et al*., 2011) to keep sites separated by at least 100 bases. The analysis was run including all possible quartets of samples and with 500 bootstrap replicates. The resulting trees were visualized with the software Figtree v1.4.3 (Rambaut, 2016) and the topologies compared with the R package phytools (Revell, 2012). I also generated unrooted neighbournets using the software SplitsTree (Huson, 1998) from the alignment file with all the loci and taxa.

Evaluating genetic structure

In order to infer the genetic structure at the whole genus level I randomly took a subsample of 31530 unlinked loci used for phylogenetic analysis using the tool *vcfrandomsample* within the vcflib package (Garrison, 2018). I also performed a principal components analysis using Tassel 5.2.31 (Bradbury *et al*., 2007) based on pairwise nuclear genome distances and a multidimensional scaling. I also performed a Bayesian analysis of the genetic structure using the software fastStructure (Raj *et al*., 2014) with values of K from 1 to 23 and three replicates for each value. I used the

python script choosek.py to identify the value of *K* that maximises the marginal likelihood.

Hybridisation analysis

To estimate the rate of hybridisation I used the python package HyDe (Blischak *et al*., 2018). HyDe works by randomly choosing three taxa from the database in order to build a phylogenetic network per site rooted on a specified outgroup taxon. Each member of this network is assigned to one of three categories, parental taxon 1 (P1), parental taxon 2 (P2) and hybrid (Hyb). Under a coalescent model with no hybridization, the produced networks can have either of two possible topologies. In the first topology, the closest relative to the hybrid taxon is the parental taxon P1 with a likelihood of 1-*y*. In the second topology the closest relative to the hybrid taxon is the parental taxon P2 with a likelihood of *y*. Therefore gamma is defined as the likelihood of the hybrid taxon being closely related to either the parental taxa 1 or the parental taxa 2 (P1 or P2). Additionally, the ratio of these topologies is used to detect hybridization and to estimate the amount of admixture between taxa.

As a first step I analysed all possible triplets (178,920) considering each taxon as a different species (73 taxa including the outgroup). Samples of *A. barrelieri* and *A. hispanicum* from Morocco were considered as different taxa for this analysis as they were clustered in different clades in the phylogenetic analyses. I used the filtered output file with significant results to conduct bootstrap resampling of individuals for the specified triplets with 200 replicates. Finally, the distribution of gamma values for the significant combinations of taxa were plotted using *ggplot2* (Wilkinson, 2011) in R.

Ancestral state reconstruction of vegetative and reproductive characters

Seven ecologically important morphological characters in *Antirrhinum* were chosen to perform ancestral character reconstruction. Three of them are continuous vegetative traits (height, branching pattern, leaf shape), three are continuous reproductive traits (nodes to flower, pedicel length and flower size) and the last one

is a discrete reproductive trait (flower colour). These characters were scored on plants from the same accessions used for phylogenetic analysis growing under greenhouse conditions.

Nodes to flower is the number of nodes in the main stem until the first flower is produced. This number was then divided by the number of nodes until the first flower in the longest axillary branch to calculate an estimate of branching as an approximation of plant habit. With this trait, plants with a high branching index have shorter branches in comparison to the main stem. Height was measured in cm as the distance from the cotyledons to the first flower.

One leaf from each node and one flower from each plant were used for further data collection. Samples were dissected and attached to a piece of paper for digitalisation using a Nikon Coolpix 4500 camera. The length of the pedicel was measured in centimetres. Because variation in leaf or flower size was correlated with variation in shape, AAMToolbox (Hanna, 2007) was used to describe variation in both by a PCA analysis. In order to better capture the relationship between size and shape on these organs I used the PCA coordinates in the ancestral reconstruction. Specifically, I used the first PC as it captures most of the correlation between shape and size.

Colour patterns of the corolla were scored based on well characterised phenotypes and genotypes previously recorded in mutants of the species *A. majus.* Flower colour was classified as white, yellow, magenta or restricted magenta. A flower was considered yellow when the colour was spread through the whole area of the petal lobes and restricted magenta when the red pigmentation was focused in the middle part of the petal lobes.

I performed maximum-likelihood ancestral state reconstruction for all characters. For the continuous traits I used the fastAnc and contMap functions in phytools (Revell, 2012) assuming a Brownian model of evolution. In the case of the categorical trait flower colour species could be considered polymorphic if different individuals within a species have different flower colours. For this reason I used two different approaches to deal with existence of polymorphic individuals. In the first approach I

used the *ace* function in Ape (Paradis *et al*., 2004) and, in order to account for the polymorphic states, I included each alternative state for each taxa in a different tip in the phylogeny. For example, I included white and magenta states for the species *A. boissieri* into two separate accessions in the phylogeny: A. boissieri_L18_2 and A.boissieri_L18_2. Then I tested three different models of evolution: 1) a symmetrical model in which all transitions between characters are possible and forward and reverse transitions have equal rates; 2) an asymmetrical model in which all transitions are possible but forward and reverse transitions have different rates and 3) a constrained model following the assumption that forward and reverse transitions between yellow and magenta and yellow and restricted magenta are not possible.

For the second approach, I used the *rayDisc* function in the package corHMM (Beaulieu, *et al*., 2013). This function performs a maximum likelihood reconstruction allowing for polymorphic characters. The way in which rayDisc deals with polymorphic data is by assigning equal values of likelihood to each state in a polymorphic sample. In other words, if sample_10 has four different states (A, B, C, D) but is polymorphic in A and B, rayDisc will assign a likelihood of 1.0 to both states. For this analysis I only included one individual per taxon in the phylogeny and tested four different models of evolution. These models were the same symmetrical and asymmetrical models tested with Ape (Paradis *et al*., 2004) but I included a constrained alternative for each one. I evaluated the degree of support for each model using a likelihood ratio test and the Akaike Information Criterion (AIC) score. To facilitate the visual analysis of the results I transformed the branch lengths of the ML likelihood tree using the function *compute.brlen* and the Grafen method from Ape, as some of the branches were extremely short in the original phylogenetic tree.

Results

Phylogenetic relationships

The phylogenetic analyses clearly resolve the major relationships within *Antirrhinum*. The ML (figure 2.3) and coalescent (figure S.1) trees shows similar overall topologies recovering the relationships between the three main taxonomic subsections with high support. In both trees all subsections are shown as paraphyletic (figure S.2). For example, the species *A. braun-blanquetii* traditionally classified within *Streptosepalum* is placed within and sister to *Kickxiella* in the ML and coalescent trees respectively. Subsection *Kickxiella* is divided into four different clades. Clade I is placed as the most basal group in the genus *Antirrhinum* in both phylogenies. Clade II is composed of the species *A. molle* only and it is placed between subsections *Antirrhinum* and *Streptosepalum*. Clade III is nested within subsection *Antirrhinum* and is composed of *A. hispanicum* from Morocco. Finally, Clade IV is also placed within subsection *Antirrhinum* and is composed of different species distributed in the Sierra Nevada (figure 2.3).

The topology of both phylogenetic trees shows a strong geographic structure with certain species grouped mainly by their distribution instead of morphology. This is the case for clade III of *Kickxiella*, which is nested within *Antirrhinum* and grouped with other species distributed in the Sierra Nevada. Also, the accessions of the species *A. barrelieri and A. hispanicum* from the Sierra Nevada are more closely related to each other than to the conspecific accessions from Morocco.

As expected, most nodes have high bootstrap support in the ML tree, but varying levels of support in the coalescent tree. This variability in support is also shown by the amount of conflict present in the unrooted neighbournet, depicted as the number of lateral connections between species (figure 2.4). Each one these lateral connections represent a different alternative topology conflicting with the genome-wide majority.

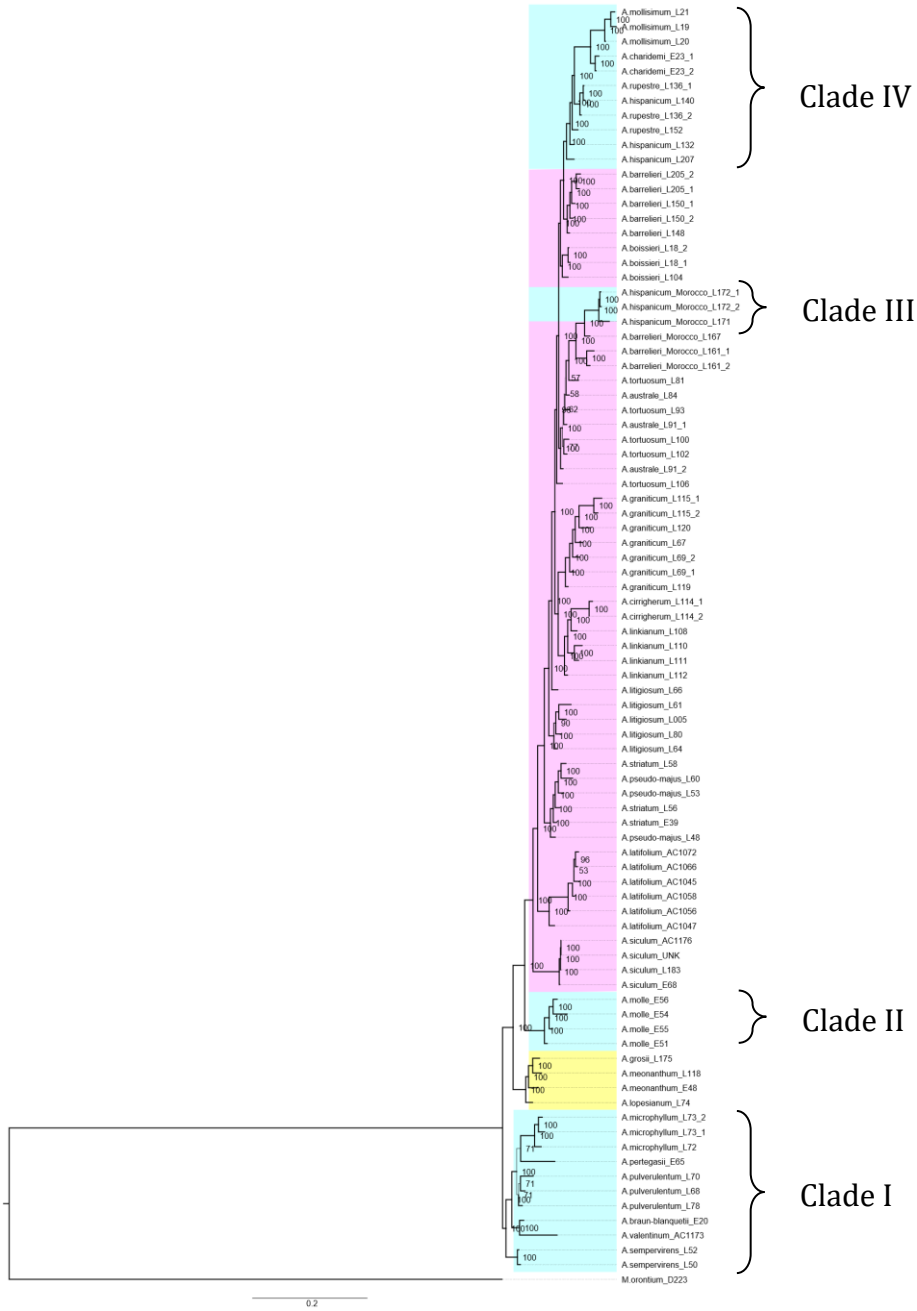


Figure 2.3. Maximum likelihood tree for 26 *Antirrhinum* taxa based on 16,061,293 aligned nucleotides analysed using RAxML with a GTR-GAMMA model of nucleotide substitution. Bootstraps values are given on the tree. The colours represent the three main subsections: blue corresponds to subsection *Kickxiella* (clades are numbered for discussion) yellow to subsection *Streptosepalum* and pink to subsection *Antirrhinum*.

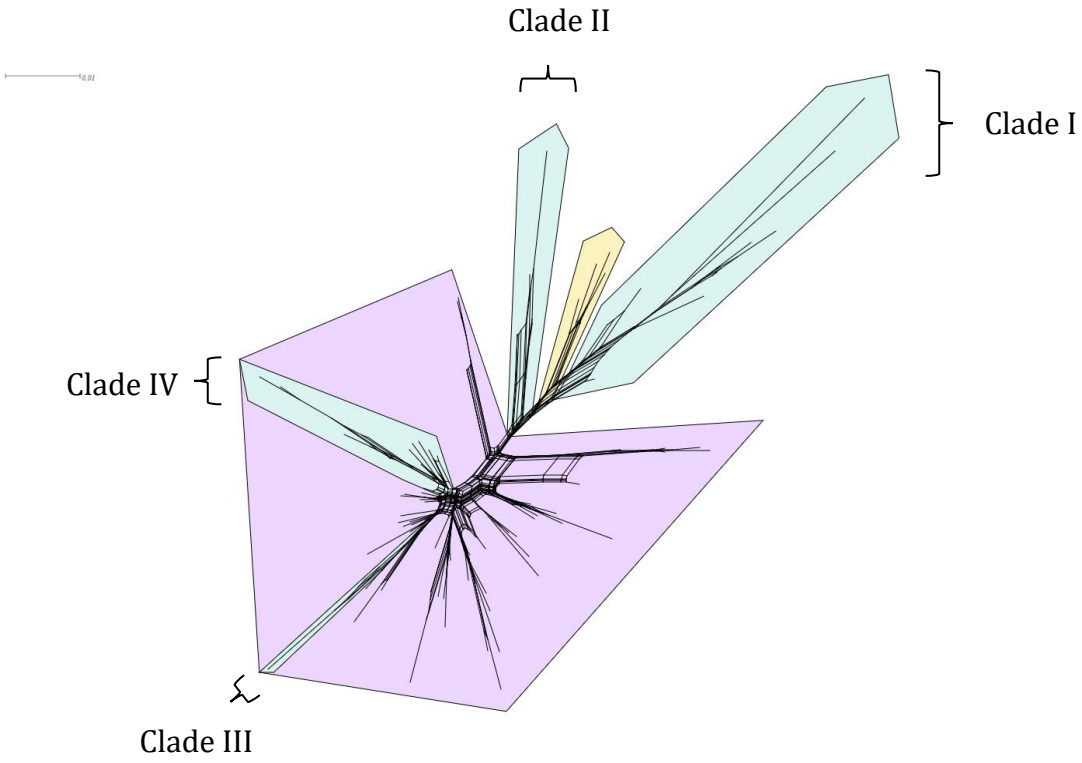


Figure 2.4. Unrooted neighbour-net graph produced with SplitsTree using RAD sequencing alignment. Lateral connections show conflicting topologies between species and subsections. Colours and labels correspond to figure 2.3.

Genetic structure

The PCoA identifies two main groups corresponding to the clades in the phylogenetic tree. The first group is composed of clades I and II from subsection *Kickxiella* and subsection *Streptosepalum*. The second group corresponds to subsection *Antirrhinum* plus clades III and IV of subsection *Kickxiella* (figure 2.5). The species *A. latifolium, A. molle*, and *A. siculum* are placed as intermediate between the two subsections. This pattern is also evident in the results from the fastSTRUCTURE analysis with the optimal K value of 2 (figure 2.6). fastSTRUCTURE also recognizes the same two groups shown in the PCoA but also shows species *A. latifolium*, *A. molle*, and *A. siculum* having equal genetic contributions from each subsection. Results from K=3 recognise species *A. latifolium* and *A. siculum* from subsection *Antirrhinum* as a distinct genetic cluster.

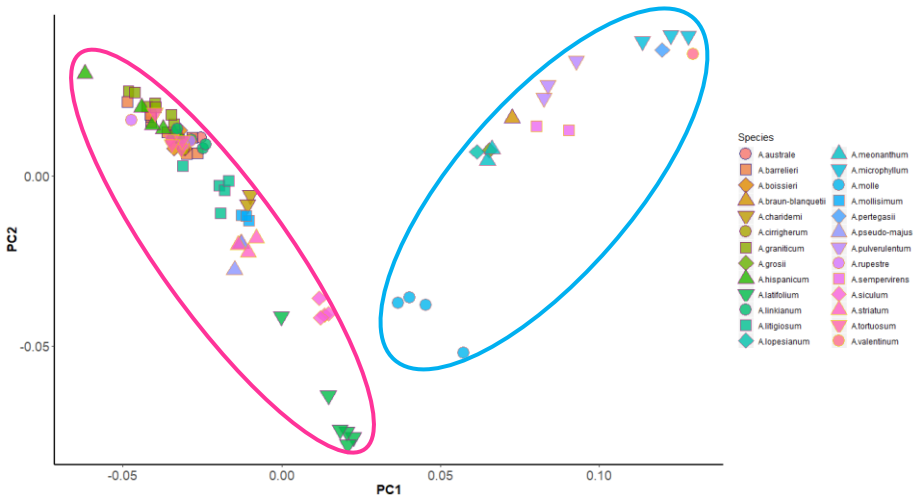


Figure 2.5. Principal coordinate analysis of the genus *Antirrhinum* elaborated using a 31530 loci. The first PC separates subsections *Kickxiella* (blue circle) and *Antirrhinum* (magenta circle).

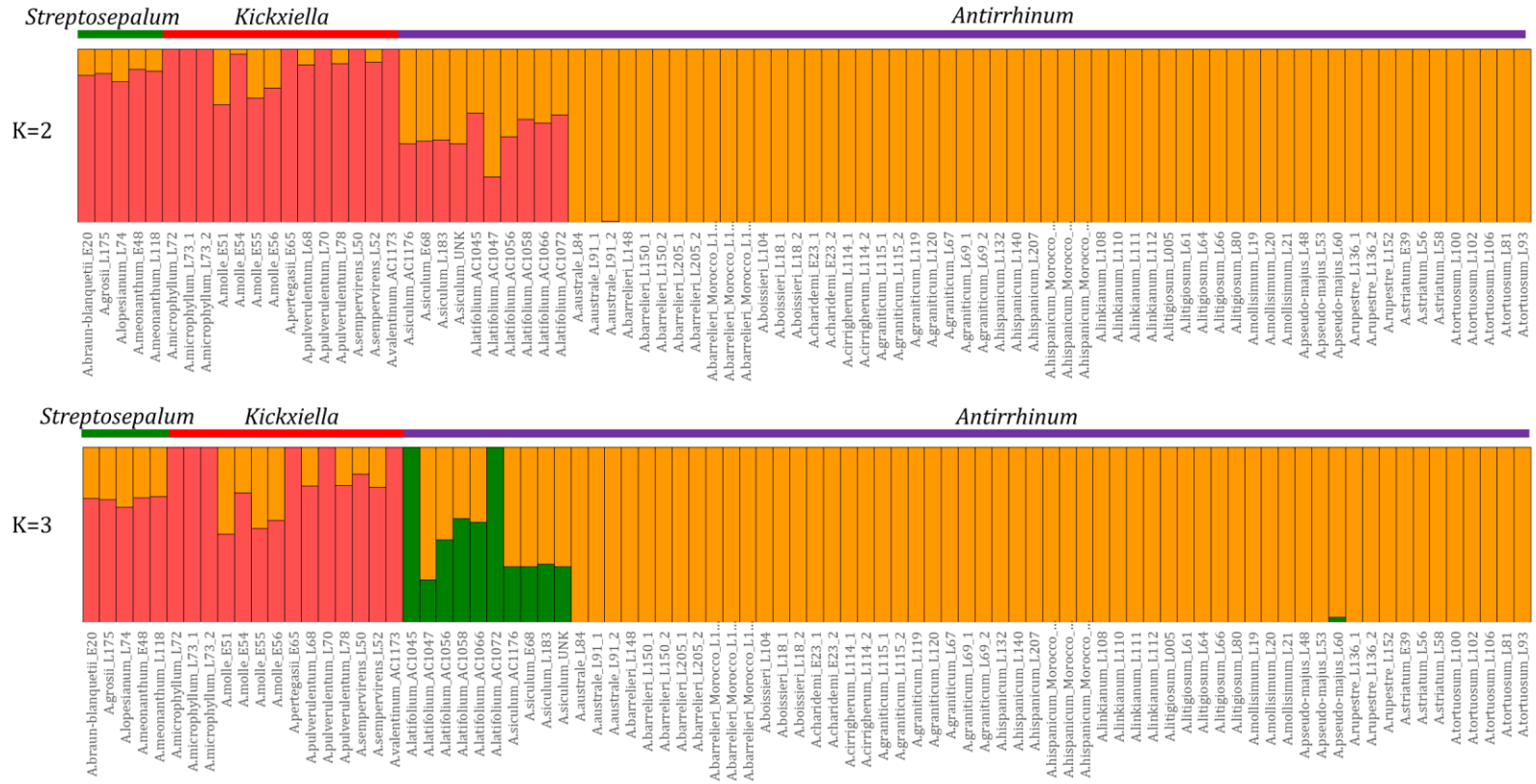


Figure 2.6. Genetic structure of the 26 species of *Antirrhinum* estimated in fastSTRUCTURE. Results are shown for K=2 (top) and K=3 (bottom). A maximized marginal likelihood comparison indicates that K=2 is the best-fitting number of groups. Labels are organised alphabetically within each genetic cluster.

Hybridisation analysis

The analysis of hybridisation with HyDe showed 48,672 of 178,920 (27.20%) significant tests (triplets) after a Bonferroni correction ($0.05/178{,}920=2.7\text{x}10^{-7}$). High levels of hybridisation occurring between all *Antirrhinum* species and populations were found with most gamma values distributed between 0.3 and 0.6 (figure 2.7).

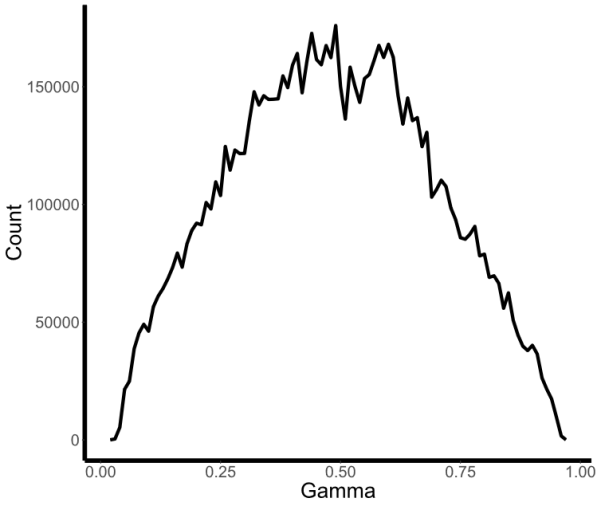
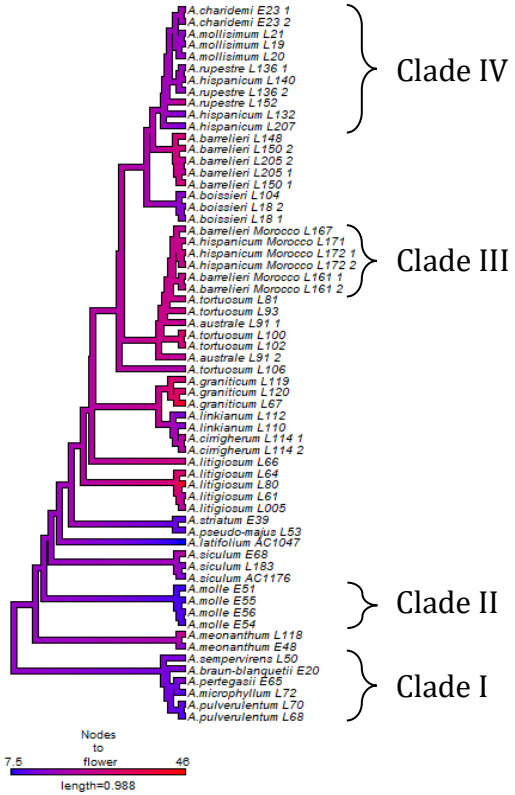
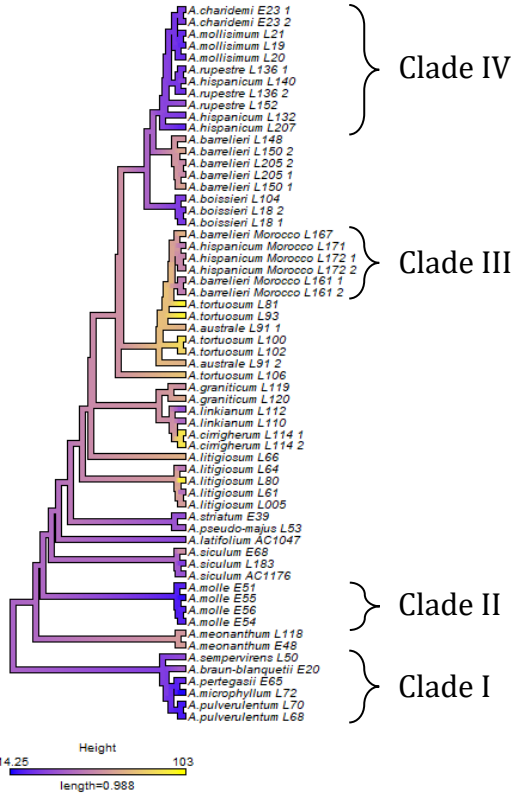


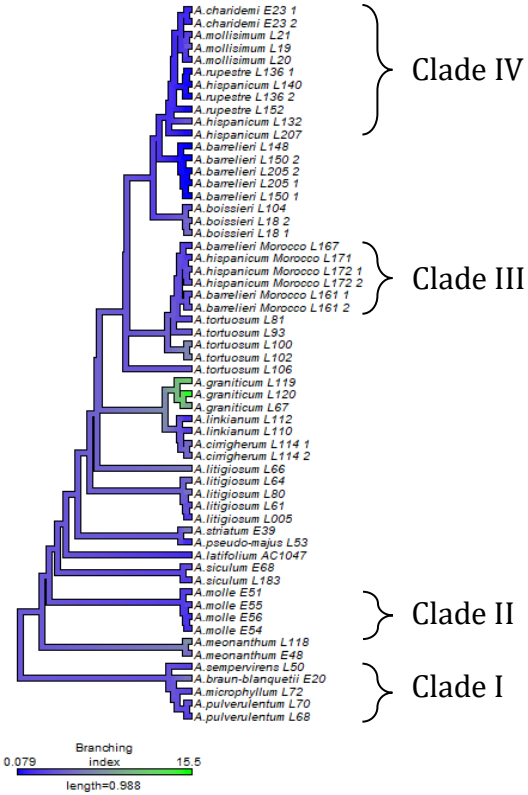
Figure 2.7. Distribution of gamma values obtained in HyDe across 200 bootstrap replicates reveals high levels of hybridisation.

Continuous vegetative and reproductive variables

The ancestral reconstruction of vegetative traits showed the ancestral *Antirrhinum* phenotype was likely short, with a prostrate habit and a fast reproductive cycle (figure 2.8). The data also showed a general tendency towards more derived taxa flowering from a later node, and growing taller with multiple transitions from a prostrate habit to an upright habit within subsection *Streptosepalum* and subsection *Antirrhinum*. Within subsection *Antirrhinum*, *Kickxiella* clade IV and the species *A. boissierii* shows traits similar to the ancestral *Antirrhinum,* but *Kickxiella* clade III presents traits similar to the rest of subsection *Antirrhinum*.


A.charidemi E23 1
A.charidemi E23 2
A.mollisimum L21
A.mollisimum L19
A.mollisimum L20
A.rupestre L136 1
A.hispanicum L140
A.rupestre L136 2
A.rupestre L152
A.hispanicum L132
A.hispanicum L207
A.barnelieri L148
A.barnelieri L150 2
A.barnelieri L205 2
A.barnelieri L205 1
A.barnelieri L150 1
A.boissieri L104
A.boissieri L18 2
A.boissieri L18 1
A.barnelieri Morocco L167
A.hispanicum Morocco L171
A.hispanicum Morocco L172 1
A.hispanicum Morocco L172 2
A.barnelieri Morocco L161 1
A.barnelieri Morocco L161 2
A.tortuosum L81
A.tortuosum L93
A.australe L91 1
A.tortuosum L100
A.tortuosum L102
A.australe L91 2
A.tortuosum L106
A.graniticum L119
A.graniticum L120
A.graniticum L67
A.linkianum L112
A.linkianum L110
A.cimigherum L114 1
A.cimigherum L114 2
A.litigiosum L66
A.litigiosum L64
A.litigiosum L80
A.litigiosum L61
A.litigiosum L005
A.striatum E39
A.pseudo-majus L53
A.latifolium AC1047
A.siculum E68
A.siculum L183
A.siculum AC1176
A.molle E51
A.molle E55
A.molle E56
A.molle E54
A.meonanthum L118
A.meonanthum E48
A.sempervirens L50
A.braun-blanquetii E20
A.pertegasii E65
A.microphyllum L72
A.pulverulentum L70
A.pulverulentum L68
Clade IV
Clade III
Clade II
Clade I
Nodes to flower
7.5
46
length=0.988


A.charidemi E23 1
A.charidemi E23 2
A.mollisimum L21
A.mollisimum L19
A.mollisimum L20
A.rupestre L136 1
A.hispanicum L140
A.rupestre L136 2
A.rupestre L152
A.hispanicum L132
A.hispanicum L207
A.barnelieri L148
A.barnelieri L150 2
A.barnelieri L205 2
A.barnelieri L205 1
A.barnelieri L150 1
A.boissieri L104
A.boissieri L18 2
A.boissieri L18 1
A.barnelieri Morocco L167
A.hispanicum Morocco L171
A.hispanicum Morocco L172 1
A.hispanicum Morocco L172 2
A.barnelieri Morocco L161 1
A.barnelieri Morocco L161 2
A.tortuosum L81
A.tortuosum L93
A.australe L91 1
A.tortuosum L100
A.tortuosum L102
A.australe L91 2
A.tortuosum L106
A.graniticum L119
A.graniticum L120
A.linkianum L112
A.linkianum L110
A.cimigherum L114 1
A.cimigherum L114 2
A.litigiosum L66
A.litigiosum L64
A.litigiosum L80
A.litigiosum L61
A.litigiosum L005
A.striatum E39
A.pseudo-majus L53
A.latifolium AC1047
A.siculum E68
A.siculum L183
A.siculum AC1176
A.molle E51
A.molle E55
A.molle E56
A.molle E54
A.meonanthum L118
A.meonanthum E48
A.sempervirens L50
A.braun-blanquetii E20
A.pertegasii E65
A.microphyllum L72
A.pulverulentum L70
A.pulverulentum L68
Clade IV
Clade III
Clade II
Clade I
Height
14.25
103
length=0.988


A.charidemi E23 1
A.charidemi E23 2
A.mollisimum L21
A.mollisimum L19
A.mollisimum L20
A.rupestre L136 1
A.hispanicum L140
A.rupestre L136 2
A.rupestre L152
A.hispanicum L132
A.hispanicum L207
A.barnelieri L148
A.barnelieri L150 2
A.barnelieri L205 2
A.barnelieri L205 1
A.barnelieri L150 1
A.boissieri L104
A.boissieri L18 2
A.boissieri L18 1
A.barnelieri Morocco L167
A.hispanicum Morocco L171
A.hispanicum Morocco L172 1
A.hispanicum Morocco L172 2
A.barnelieri Morocco L161 1
A.barnelieri Morocco L161 2
A.tortuosum L81
A.tortuosum L93
A.tortuosum L100
A.tortuosum L102
A.tortuosum L106
A.graniticum L119
A.graniticum L120
A.graniticum L67
A.linkianum L112
A.linkianum L110
A.cimigherum L114 1
A.cimigherum L114 2
A.litigiosum L66
A.litigiosum L64
A.litigiosum L80
A.litigiosum L61
A.litigiosum L005
A.striatum E39
A.pseudo-majus L53
A.latifolium AC1047
A.siculum E68
A.siculum L183
A.molle E51
A.molle E55
A.molle E56
A.molle E54
A.meonanthum L118
A.meonanthum E48
A.sempervirens L50
A.braun-blanquetii E20
A.microphyllum L72
A.pulverulentum L70
A.pulverulentum L68
Clade IV
Clade III
Clade II
Clade I
Branching index
0.079
15.5
length=0.988

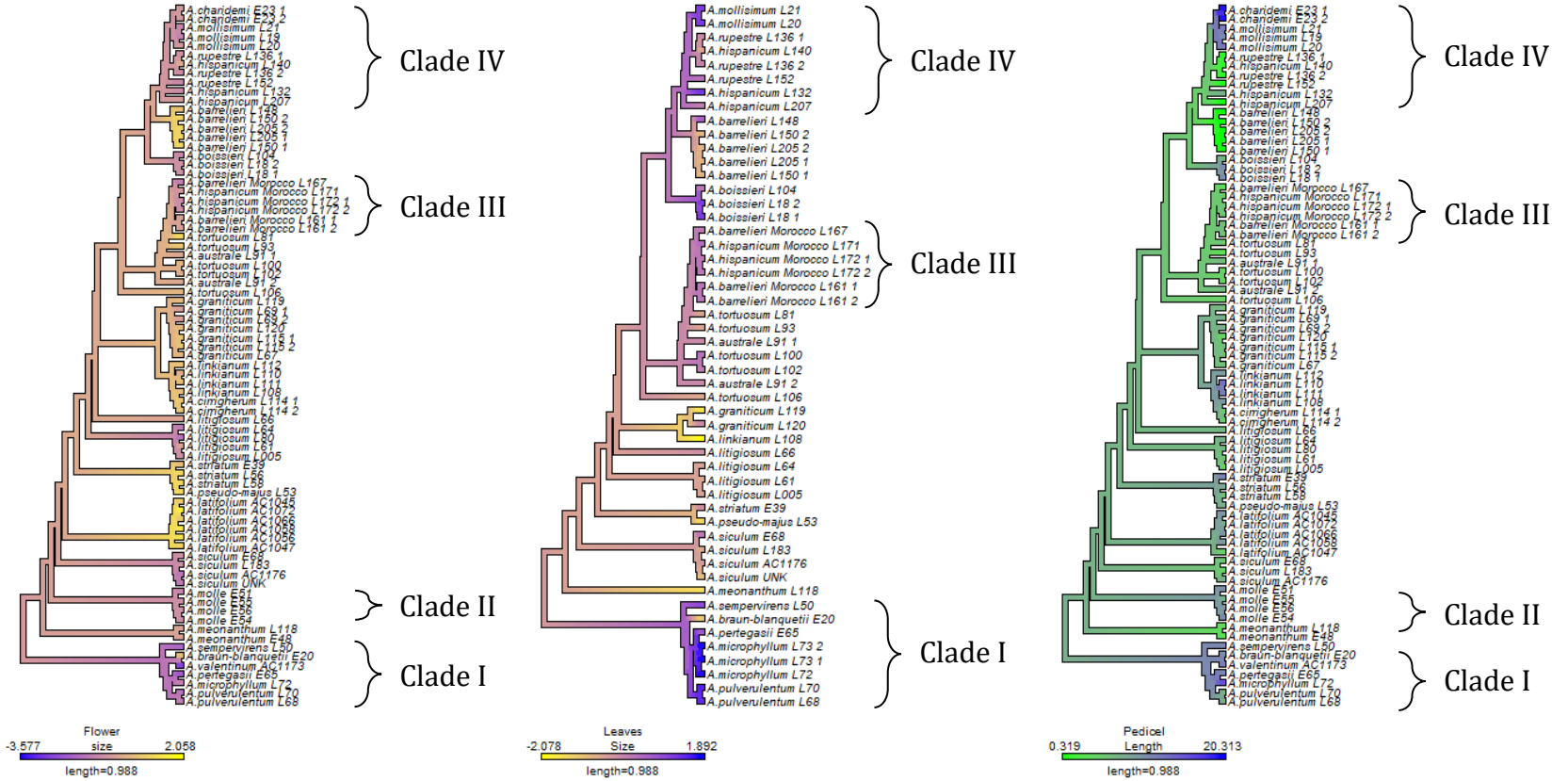


Figure 2.8. Ancestral state reconstruction of continuous characters with fastAnc and contMap functions in phytools (Revell, 2012) assuming a Browninan model of evolution. Labels correspond to figure 2.3.

Flower colour

There are not significant differences between symmetric, asymmetric or constrained models in any of the two approaches followed (table 1) but the symmetrical model was chosen as it has the lowest AIC value in both approaches with no significant differences between constrained and unconstrained. Here I show the results obtained with corHMM (Beaulieu, *et al.*, 2013) as it deals better with polymorphism in the data.

White or yellow are recovered as ancestral for *Antirrhinum* flowers with multiple independent transitions from white to yellow in each subsection (figure 2.9). Magenta pigmentation becomes predominant in *Antirrhinum* after a transition from yellow to magenta observed in the clade formed by the taxa *A. striatum* and *A. pseudomajus* with several reverse transitions to white throughout the subsection. These reverse transitions to white flowers are especially common in taxa characterised for having flowers with restricted magenta pigmentation (i.e. A. *graniticum*, *A. mollisimum*, *A. boisseri* and the Moroccan *A. hispanicum*).

The maximum likelihood estimates of the transitions rates under the symmetrical model support a constrained evolution of flower colour in *Antirrhinum* (table 2.1). Forward and reverse transitions between yellow and magenta (restricted or not) have a likelihood of zero (table 2.2). Additionally, forward and reverse transitions from white to yellow are more likely to occur than transitions from white to magenta (restricted or not).

Table 2.1. Comparison of models tested in this study using the R packages Ape and corHMM.

Ape		
Model	**AIC**	**Lnlik**
Symmetrical	181.09	-84.54
Asymmetrical	188.25	-82.12
Constrained	193.9	-88.95
corHMM		
Model	**AIC**	**Lnlik**
Symmetrical	70.68	-29.34
Symmetrical constrained	72.5	-28.25
Asymmetrical	80.49	-28.24
Asymmetrical constrained	80.41	-28.2

Table 2.2. Maximum likelihood estimates of transition rates under the symmetrical model obtained in corHMM.

	White	**Yellow**	**Magenta**	**Restricted Magenta**
White	-	100.00	0.12	1.06
Yellow	100.00	-	0.00	0.00
Magenta	0.12	0.00	-	14.89
Restricted Magenta	1.01	0.00	14.89	-

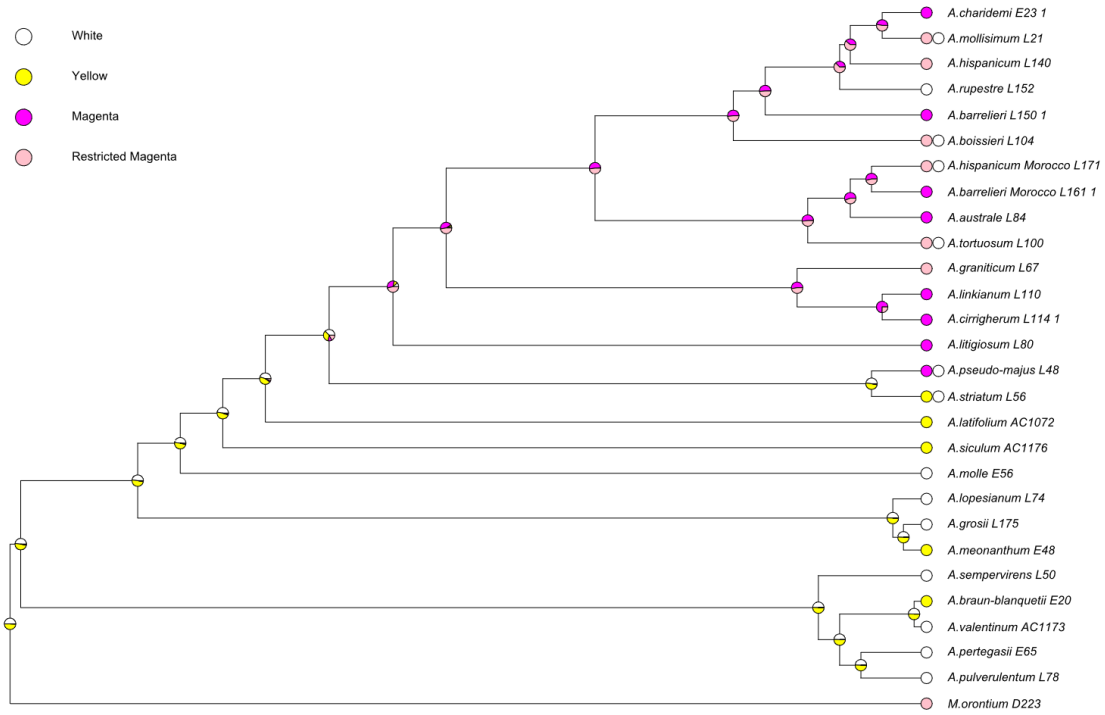


Figure 2.9. Ancestral reconstruction of flower colour estimated with corHMM under a symmetrical model. Pie charts on nodes represent the likelihoods of each state. Coloured circles at the tips represent species' states with polymorphic species having two different circles.

Discussion

The results obtained in this chapter show the utility of RAD sequencing as a powerful approach for the phylogenetic reconstruction of recently radiated taxa. A phylogenetic analysis of the genus *Antirrhinum* and its association with phenotypic traits shows paraphyletic morphological subsections, suggesting that the distinct suite of morphological traits of subsection *Kickxiella* have arisen multiple times via parallel phenotypic evolution. While the genetic underpinnings of this parallel trait evolution are currently unclear, my results also suggest that hybridisation has played a major role in the evolution of species distributed within the same geographic region opening a new set of questions about the role of introgression and the exchange of adaptive traits in *Antirrhinum*. I discuss the evidence supporting these results and the implication for understanding the species radiation and morphological divergence in *Antirrhinum*.

Phylogenetic relationships

A phylogenetic analysis shows the multiple occurrence of subsection *Antirrhinum*, *Kickxiella* and *Streptosepalum* with some individuals grouped based on morphology and others on geography. For example, the only accession of the species *A. braun-blanquetii*, traditionally classified within subsection *Streptosepalum*, was placed in subsection *Kickxiella* in both trees. The morphological analysis shows that this individual has similar vegetative traits to subsection *Kickxiella* and the results of genetic structure suggest a possible hybrid origin, explaining its placement in this subsection. The results also show four clades of species with *Kickxiella*-like morphologies distributed across the phylogeny, two of them within subsection *Antirrhinum*. Two of these clades (I and III) were found previously using a neighbour joining analysis of AFLP markers (Wilson and Hudson, 2011). In this earlier study, Wilson and Hudson (2011) name clade I as "core *Kickxiella*" mainly because their shared phenotype and their non-overlapping distributions suggest these species descend from a common ancestral *Kickxiella* species. They also report that accessions of *A. hispanicum* from Morocco (clade III) and from the Sierra Nevada

have different morphologies, with the Moroccan accessions having an *Antirrhinum* morphology and Spanish accessions having a *Kickxiella* morphology. The same results were found in this work and therefore clade III will not be considered as part of subsection *Kickxiella*. This incongruence between morphology, taxonomy and genetic relationships explains why morphological characters have not been useful in the past to define genetic relationships in the genus (Adeyanju, 2003; Chaffe, 2003; Vargas *et al*., 2004) and suggest the necessity of a new taxonomic revision based on morphology and genetic data.

The multiple appearance of *Kickxiella* morphologies across the phylogenetic tree can be explained by the interplay of different evolutionary forces acting at different times during *Antirrhinum* speciation. Here I propose three different hypothesis that could lead to the topology observed, discuss the evidence supporting each and explain their limitations.

In a first scenario, an initial radiation event in the genus could have been followed by multiple events of parallel evolution of *Kickxiella* morphologies. Under this scenario, the development of *Kickxiella* morphology would come from the origin of new mutations arising independently in each clade generating similar morphologies and caused by the occurrence of similar selective pressures. In this work, the phylogenetic clustering of certain species by geographic proximity and not by morphology is consistent with this hypothesis. The process of a parallel evolution of adaptive traits within a single radiation event has been documented in several lineages of plants and animals. For example, in stickleback fish the recurrent loss of pelvic spines in fresh water species is caused by a reduction in the predation regime in comparison to the salty water habitats (Chan *et al*., 2010). In the *Senecio lautus* complex, multiple events of adaptation to coastal environments occurred independently after an initial radiation event (Roda *et al*., 2013). In *Antirrhinum* we cannot rule out this possibility. However, this scenario seems unlikely as a rapid and recent radiation means a low chance for the evolution of complex morphological traits to occur multiple times. Despite the wide variation in morphology observed

between populations, all *Kickxiella* species are characterized by having a very similar set of adaptive traits. Some of these characters have been found to be controlled by multiple underlying genes in experimental crosses between the species *A. charidemi* (Subsection *Kickxiella*) and *A. majus* (Subsection *Antirrhinum*). Altogether, this makes it unlikely that there would be morphological convergence in independent lineages within a single radiation event (Langlade *et al*., 2005; Feng *et al*., 2009).

A second scenario that could explain the topology of the tree is the occurrence of an initial adaptive radiation in *Antirrhinum* followed by multiple events of hybridisation and the lateral transfer of adaptive alleles. Ancient hybridisation events have been considered as an important factor in the evolution of the genus with multiple evidence supporting its occurrence. For example, by looking at the distribution of important morphological characters in *Antirrhinum*, Rothmaler (1956) and Webb (1972) proposed a model of isolation-contact-isolation for the evolution of the genus, caused by climatic fluctuations during the last glaciation. These multiple events of secondary contact after long periods of isolation would open the possibility for the introgression of adaptive alleles from one subsection to the other, allowing the colonisation of new habitats. Evidence for hybridisation has also been found in previous works using nuclear and plastid markers. For example, Vargas *et al.* (2004) and Vargas *et al.* (2009) report the occurrence of nucleotide additivity (two or more bases detected at the same nucleotide position) in the ITS region, supporting recent hybridisation among species. Vargas *et al.* (2009) found evidence for species within the same broad geographic area sharing chloroplast haplotypes. Likewise, Wilson & Hudson (2011) found a mismatch between chloroplast lineages and morphology in species distributed in the same geographical regions. Here we found support for this scenario in the rampant levels of hybridisation at the whole genus level suggested by the analysis with HyDe.

Finally, another scenario that could explain the distribution of *Kickxiella* morphologies across the tree is based on the fact that the two previous hypotheses are not mutually exclusive. In other words, the observed pattern of morphological

evolution in *Antirrhinum* could also be driven by the processes of hybridisation and parallel evolution occurring in different clades in the phylogeny. Firstly, we have evidence of hybridisation in the results from the PCoA and fastSTRUCTURE analyses. Both results split the genus into two groups that mainly correspond to subsection *Kickxiella* clades I and II plus *Streptosepalum*, and subsection *Antirrhinum* plus *Kickxiella* clades III and IV. However, the species *A. latifolium*, *A. molle* and *A. siculum* were placed between these groups showing an equal estimated proportion of ancestry from each side. In addition, most taxa within the *Kickxiella* clades I and II show evidence of admix ancestry in the results from fastSTRUCTURE. Secondly, species in the *Kickxiella* clade IV do not show any degree of introgression from the other *Kickxiella* clades, therefore suggesting independent morphological evolution and supporting the idea of parallel evolution occurring in subsection *Antirrhinum*.

Ancestral character reconstruction

The ancestral *Antirrhinum* phenotype was estimated to be short, with a prostrate habit, a fast reproductive cycle and white or yellow flowers. This corresponds with the phenotype observed in the members of subsection *Kickxiella* (clade I, II and IV). Additionally, the results show that all the *Kickxiella* clades have a very similar morphology without much variation. On the contrary, species traditionally classified as *Antirrhinum* have a wide range of morphological variation, although there is a tendency to be taller, with longer leaves and bigger flowers. This phenotype seems to have evolved independently in subsection *Streptosepalum*, supporting the previous hypothesis of parallel evolution. The same pattern can be observed in the distribution of flower colours in the phylogeny. Subsection *Antirrhinum* is the clade with the greatest variation in flower colour and the largest number of taxa.

In this work we could not find any differences between any of the models tested independently of the treatment given to polymorphic data. A possible reason is that the power of this kind of comparative methods rely in the amount of transitions observed between character states in a given phylogeny (Omland, 1999). Therefore, it is possible that the *Antirrhinum* phylogeny does not have enough number of

transitions between colours in order to test different models of evolution with enough confidence. Although this means we should consider the results obtained here with caution, previous works in *Antirrhinae* (Ellis and Field, 2016) and other model systems like *Solanum* (Ng *et al*., 2018) support the results found in this study. In the future, more complex models should be tested accounting for other variables such as differences in diversification rates between different clades in the phylogeny, e.g. those implemented in the software diversitree (Fitzjohn, 2012).

My results suggest that the radiation in subsection *Antirrhinum* was likely to be accompanied by a shift in flower colour from white to magenta, suggesting an adaptive role of flower colour in speciation. The influence of colour in the evolution of *Antirrhinum* has been tested in previous studies. Whibley *et al*. (2006) compared the spatial distribution of loci controlling yellow and magenta colours against neutral loci in a hybrid zone between the taxa *A. pseudomajus* and *A. striatum*. They concluded that the distribution of the loci controlling flower colour in the hybrid population has been shaped by natural selection maintaining reproductive isolation between these taxa. In addition, other studies have found an effect of floral enhancers like venation and pigmentation patterns on pollinator preference, supporting the role of floral morphology in the speciation of *Antirrhinum* (Shang *et al*., 2011; Whitney *et al*., 2013; Bradley *et al*., 2017).

The pattern of flower colour variation in the genus can be explained by evolutionary constrains in the biosynthetic pathway. Clades at the base of the tree have predominantly white flowers with multiple independent transitions to yellow. After a transition from white to magenta, magenta becomes common in the genus followed by multiple transitions back to white. This pattern in the evolution of colour in the phylogeny matches constrains imposed by the anthocyanin biosynthetic pathway, e.g. transitions between magenta and yellow have to be via white (Ellis and Field, 2016). It has even been suggested that the structural constrains imposed by this pathway can results in a predictable trajectory towards the convergence of phenotypes in different species (Larter *et al*., 2018). The evolutionary consequences

of these constrains have been explored previously in *Antirrhinae* (Smith and Goldberg, 2015; Ellis and Field, 2016), in *Antirrhinum* (Bradley *et al*., 2017) and in other plant groups like Solanaceae (Ng *et al*., 2018). For example, Smith and Goldberg report a bias towards gaining pigmentation in the flowers in the *Antirrhinae* tribe. This result is supported by Ellis and Field (2016), who find the same bias towards the gain of yellow and anthocyanin pigmentation. Additionally, they find support for a stepwise model of evolution in *Antirrhinae* with transitions between yellow and anthocyanin pigmentation via white. The results from this thesis support this mechanism of evolution by showing that a symmetrical but constrained model is the best fit for the evolution of flower colour in *Antirrhinum* with forward and reverse transitions between magenta and yellow having a likelihood of zero.

Conclusion

The evolutionary history of the genus *Antirrhinum* shows evidence of parallel evolution. The association of phenotypic traits in the phylogeny shows that different adaptive phenotypes have evolved multiple times in *Antirrhinum* with several shifts back to the ancestral morphology. The analysis of the genetic structure also suggests hybridisation as an important force driving the evolution of the genus, raising questions about the role of introgression in *Antirrhinum* speciation. Finally, the ancestral reconstruction of morphological and reproductive traits suggest an important role of flower colour in the radiation of the genus. I suggest that patterns of evolution in flower colour can represent constrains in the biochemical pathway which can lead to phenotypic convergence in *Antirrhinum*.

Chapter 3

Population genomics of ecologically divergent *Antirrhinum* species

Population genomics of ecologically divergent *Antirrhinum* species

Abstract

Despite being considered a homogenising force in evolution, hybridisation can also be a source of new genetic and morphological variation. The study of hybridisation and especially of natural hybrid populations can increase our understanding of the mechanism involved in the maintenance of species boundaries and the evolution of reproductive barriers. Here, I used restriction site associated DNA markers to explore the genetic composition of several putative hybrid populations between the species *A. barrelieri* and *A. rupestre.* Results show a high proportion of individuals in the south of the Sierra Nevada Mountains are hybrids and suggests continuous gene flow between species. The morphology of these hybrids closely resembles the morphology of one of the parental taxa pointing to natural selection eliminating individuals with intermediate morphology from the populations. I also explored the genetic structure of several populations from five different species in the south of Spain and show the importance of the Sierra Nevada as a barrier for gene flow between populations. In this chapter I show the importance of geographic barriers and natural selection in the origin of genetic and morphologic divergence in natural populations.

Introduction

Hybridisation is recognized as a central force in the evolution of plants and animals (Mallet, 2007). In particular, it is widely accepted that a significant proportion of speciation events in plants have involved hybridisation (Soltis and Soltis, 2009). Whitney *et al*. (2010) estimated in a meta-analysis that from a total 282 plant families and 3212 genera from North America, Europe and Australia, around 40% of families and 16% of genera are involved in hybridisation in nature.

Hybridisation can happen in a wide variety of spatial and temporal conditions. Sometimes it can occur within a specific area with a clinal pattern of morphological variation between parental taxa (Heywood, 1986; Whibley *et al*., 2006). Other times it can occur several times at different time scales during short periods of secondary contact between previously isolated taxa (Abbott *et al*., 2016). Additionally, hybridisation can also be the consequence of human disturbance creating appropriate ecological conditions for the co-occurrence of species with different ecological preferences and for the establishment of the resulting admixed progeny (van Hengstum *et al*., 2012).

During hybridisation, the genetic differences accumulated during isolation interact to create novel genetic combinations in the offspring. These genetic interactions can lead to alterations in gene expression, chromosomal structure and genome size (e.g. allopolyploidy and autopoloploidy speciation). However, the evolutionary outcome of hybridisation will depend on factors like the frequency of hybridisation events, the fitness of the new hybrid progeny, and the interaction with the environment (Baack and Rieseberg, 2007). In some circumstances, hybridisation can lead to the loss of species barriers and the homogenization of populations. In others, hybridisation can result to the creation of new adaptive genotypes or reproductive barriers, leading to speciation [e.g. hybridisation in sunflowers (Rieseberg *et al*., 1995)].

A main issue in the study of hybridisation is the difficulty in detecting and measuring its effects. For example, hybridisation and incomplete lineage sorting can result in

the sharing of alleles between lineages, making them difficult to distinguish in phylogenetic analyses (Abbott *et al*., 2016). Similarly, morphological analyses are not always effective in detecting hybridisation as new genotypic combinations do not always result in intermediate phenotypes in the progeny. This is the case in transgressive segregation, where individuals may greatly exceed the parental phenotypic values (Hersch-Green and Cronn, 2009). Traditionally, these problems in detecting later-generation hybrids have been addressed by the use of both plastid markers, which are inherited only through the female in most species, and nuclear markers inherited by both parents. However, recent technological advances have revolutionised the study of hybridisation. New generation sequencing technologies allow the discovery of enough markers for the dissection of hybrid genomes, the detection of patterns of gene flow and selection, and the identification of intrinsic reproductive barriers (Hegarty and Hiscock, 2005).

These methods are particularly effective when used in the study of hybridisation in natural populations (Rieseberg *et al*., 2000). Wild hybrid populations have been considered natural laboratories for studies of evolution and speciation, as they contain a wide variety of genotypes that are the consequence of several generations of recombination and selection (Hewitt, 1988; Minder *et al*., 2007). The study of hybrid populations allow patterns of genetic variation to be related to morphological, ecological and geographical variation and therefore the exploration of different evolutionary forces in the maintenance of species boundaries (Noor and Feder, 2006). In this work I characterize the genetic structure of an overlap zone between two species capable of hybridisation to explore the mechanisms involved in the maintenance of species boundaries and the evolution of reproductive barriers between species. My main goal is to understand more about the importance of hybridisation on the speciation process.

The genus *Antirrhinum* is a plant model system for the study of developmental processes in plants and recently has emerged as a promising system for the study of factors affecting speciation in natural populations. In a previous phylogenetic

analysis (chapter 2) I identified a clade composed of five species with morphologies belonging to different taxonomic subsections distributed in the Sierra Nevada (Spain) and surrounding areas. The species *A. charidemi*, *A. hispanicum* and *A. rupestre* belong to subsection *Kickxiella*, which is characterised by a prostrate habit, small ovate leaves and axillary white flowers. The species *A. barrelieri* and *A. tortuosum* belong to subsection *Antirrhinum*, which is characterised for having an erect habit with long thin leaves and a terminal inflorescence with magenta flowers. These species overlap in range, are interfertile and are obligate outcrossers, therefore the possibility of hybridisation is very likely.

A previous study by Wilson *et al*. (Unpublished) related these different sets of morphological characters to important ecological variables. They found that individuals with *Kickxiella* morphologies prefer sites with low vegetation cover and rocky surfaces, while individuals with *Antirrhinum* morphologies occur on sites with deeper soils surrounded by a higher vegetation cover. This habitat preference suggest a role of natural selection in the maintenance of species boundaries in *Antirrhinum*.

In this work I use restriction site associated DNA markers to 1) investigate the extend of hybridisation in putative hybrid populations in the south of the Sierra Nevada between the species *A. barrelieri* and *A. rupestre* and 2) to explore the genetic structure of natural populations distributed in the south of Spain. My aim is to quantify the incidence of hybridisation in this region and to identify important barriers to gene flow between populations.

Methods

Specimen sampling and sequencing

Samples from 116 individuals from 26 populations of the species *A. barrelieri. A. rupestre*, *A. hispanicum A. mollisimum* and *A. tortuosum* were collected by A. Hudson and Y. Wilson in 2007 and 2009 in the Sierra Nevada and adjacent regions. For each population, samples were taken at least 3 m apart from each other to minimize sampling siblings. The geographic coordinates of each population are shown in table S.2.

DNA extraction and library preparation were performed as specified in chapter 2, but using *Sbf* I instead of *Pst I* as the restriction enzyme. Libraries were sequenced at Edinburgh Genomics on a HiSeq-4000 producing 164,271,806 paired end sequencing reads.

Processing and alignment of RAD-seq data

Raw reads were demultiplexed using the script *process_radtags* from the software Stacks (Catchen *et al*., 2013). The total number of raw reads sequenced per sample are shown in table S.4. Adaptors were removed and reads trimmed and filtered by quality using trimmomatic 0.36 (Bolger *et al*., 2014). Remaining reads were mapped to the version 2 *A. majus* genome using Bowtie2 (Langmead and Salzberg, 2012) and duplicated sequences removed using Picard tools(Broad Institute, 2018). Variant/invariant genotypes were called using samtools 1.6 and the multiallelic caller implemented in bcftools 1.4 (Li, 2011). This data set was then filtered by mapping quality, depth and missing data, where individuals with more than 70% of missing data and loci with less than 3 of depth, 40 of mapping quality and more than 50% of missing data were removed. After filtering, the resulting dataset was composed of 1,881,090 sites from 69 samples corresponding to 23 populations (figure 3.1).

Hybridisation

For this analysis I used only the species *A. barrelieri* and *A. rupestre* distributed in the south of the Sierra Nevada, as previous evidence supports low divergence between these taxa and suggest hybridization as a possible cause (Wilson and Hudson, 2011; Chapter 2). In order to estimate the percentage of admixed ancestry in individuals from populations in the south of the Sierra Nevada I ran STRUCTURE to estimate the number of genetically distinct populations (K) that maximises the likelihood under linkage equilibrium and Hardy-Weinberg equilibrium. Following the approach described in Smith, *et al.* (2014), I used values of K=2 to generate Q values of membership to one of two species. An individual was considered a hybrid if the Q value ranged between 0.05 and 0.95. As there was not any genetic differentiation detected in the analyses between the species *A. rupestre* and *A. hispanicum* from the south of the Sierra Nevada, I decided to consider them as a single taxonomic unit. This decision is also supported by the phylogenetic analysis in chapter 2. Hence, for this analysis I used the taxa *A. barrelieri* and *A. rupestre*/*A. hispanicum* as parental species. Also, from this point onwards I will use the species names *A. rupestre* to refer to populations of both species, *A. rupestre* and *A. hispanicum*, distributed in the south of the Sierra Nevada. STRUCTURE settings were the same as for the analyses of the population structure in chapter 2.

Evaluating population structure

A principal coordinate analysis (PCoA) was performed using Tassel 5.2.31 (Bradbury et al., 2007). The PCoA analysis was based on the calculation of a distance matrix using only sites with non-missing data for each taxon. Because I detected a relationship between geography and the position of the samples in the PC2, I performed a linear mixed effect model in *R* software v 3.5.0 (R Development Core Team, 2013) to determine if latitude and longitude has a strong effect on PC2 values. I fitted as a response variable the eigenvalues of PC2, generated from the PCoA analysis. Latitude and longitude of each taxa was fitted as a fixed effect and taxa and

populations were fitted as random effects. The model was performed using the *lmer* function of the *lme4* package (Bates *et al.*, 2015).

In order to infer the extent of individual and population admixture, I used the software STRUCTURE 2.3.4 (Pritchard et al., 2000). As I was mainly interested in broad scale genetic clustering I tested values of K from 1 to 10 running 20 independent replicates at each value of K. I used a burnin of 10^5 and a run length of 20×10^4 Markov chain Monte Carlo steps under a linkage model. The results were analysed following the Evanno *et al* (2005) method choosing the result with the highest delta K value. The optimal value of K was plotted on the geographic map of populations using the R package ggmap (Kahle and Wickham, 2013) for interpretation.

Results

Hybridisation

Results from the STRUCTURE analysis between *A. barrelieri* and *A. rupestre* populations from the Sierra Nevada revealed the presence of a substantial proportion of introgressed hybrids (figure 3.4). From a total of 57 individuals only two are classified as *A. barrelieri* and only three as *A. rupestre* with the remaining 52 individuals being classified as hybrids. Additionally, from these 52 hybrids, individuals with intermediate values of ancestry are not common in the populations with only two individuals found having values between 0.4 and 0.6 (figure 3.4).

Population structure

The principal coordinate analysis of the 69 samples clusters populations according to geography and morphology (figure 3.1). PC1 captures the morphological variation between subsections. I found a strong positive relationship between longitude and PC2 values and a negative relationship between latitude and PC2 values (table 3.1). This effect is reflected in the fact that PC2 splits populations according to their positions relative to the Sierra Nevada into north and south.

The log likelihood estimated by STRUCTURE reveals K=3 as the optimal number of genetic clusters to describe the population structure (figure 3.2), while K=2-4 also gave informative insights into population structure. Under the assumption of only two genetic clusters, STRUCTURE groups the populations according to their position relative to the Sierra Nevada into northern and southern populations. For K=3, STRUCTURE further splits the populations from the south in to two genetic cluster according to morphology (*Kickxiella* and *Antirrhinum*). Finally, with a K value of 4, STRUCTURE splits each geographical cluster into morphological groups (*Kickxiella* vs *Antirrhinum*).

Under the optimal number of genetic clusters found in STRUCTURE (K=3), the results show a discontinuous pattern of hybridisation between populations (figure 3.3). The levels of admixture between northern populations of *A. hispanicum* and southern populations of *A. barrelieri* increase the closer they are to each other, although the Sierra Nevada seems to be acting as a barrier to gene flow between populations.

Table 3.1. Coefficients of the effect of latitude and longitude on the eigenvalues from PC2. Coefficients are the result of a linear mixed effect model.

Fixed effects	**Estimate**	**S.E.**	**z**	***p***
Intercept	0.354	0.065	5.50	<0.001
Latitude	-0.009	0.002	-5.45	<0.001
Longitude	0.002	0.001	2.79	0.008
Random effects	**Variance**	**S.D.**		
Taxa	6.4e-07	0.001		
Populations	4.5e-06	0.002		

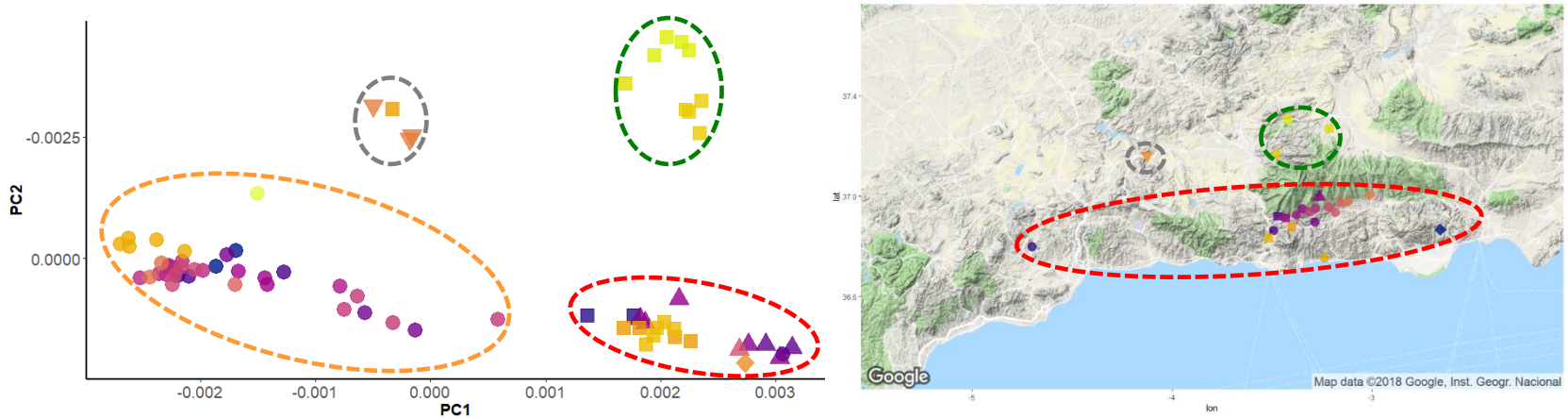


Figure 3.1. Comparison between the results from the principal coordinate analysis (left) and the geographical distribution of the 69 populations sampled (right). In both figures triangles represent *A. rupestre*, squares correspond to *A. hispanicum*, circles to *A. barrelieri*, and inverted triangles represent *A. tortuosum*. Symbols in PCA and the map are coloured by populations.

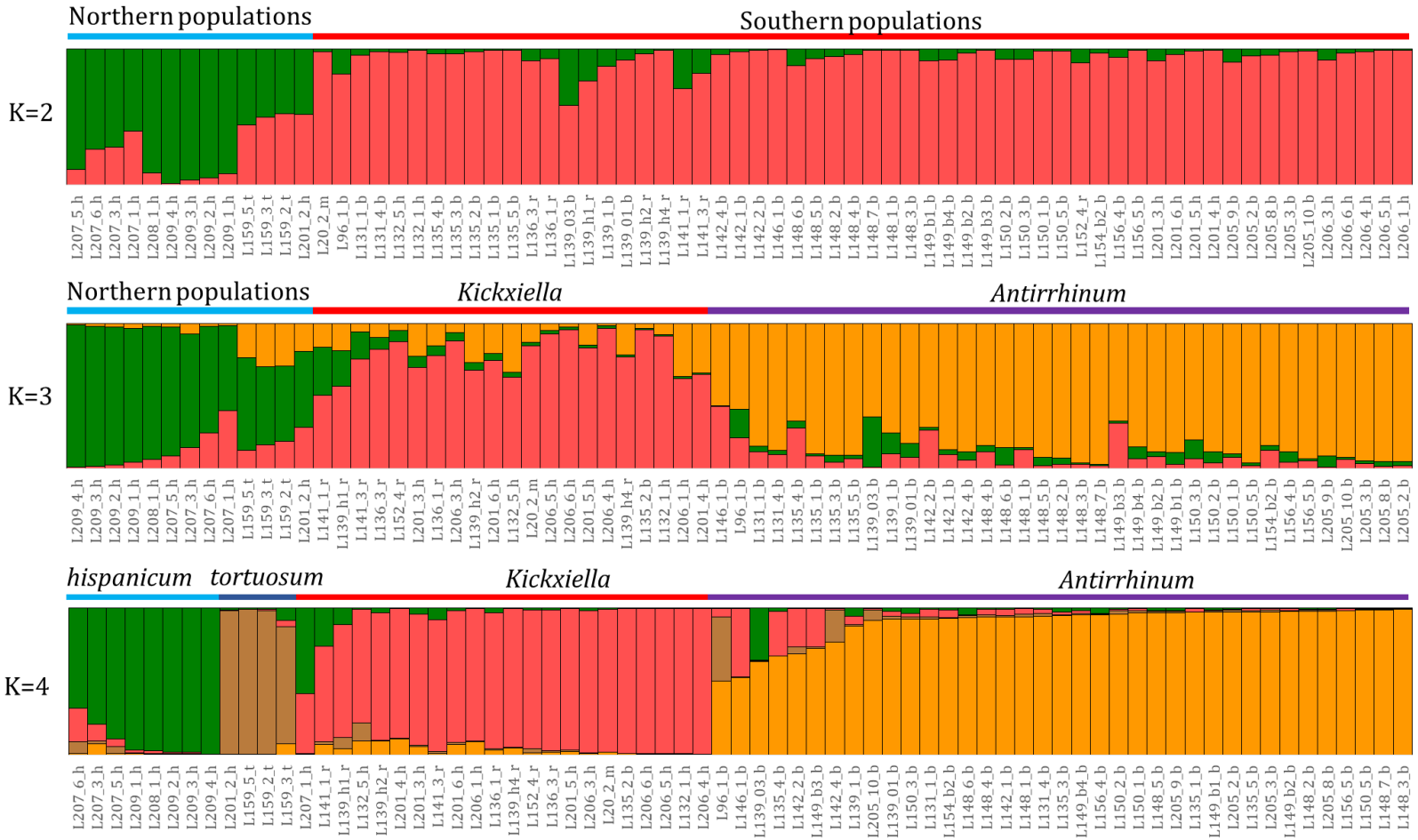


Figure 3.2. Genetic structure of the 69 samples from 23 populations from the Sierra Nevada estimated in STRUCTURE. Results are shown for K=2, K=3 and K=4. A maximized marginal likelihood comparison indicates that K=3 is the best number of groups for the genus. Colour bars on top represent the groups formed in each result.

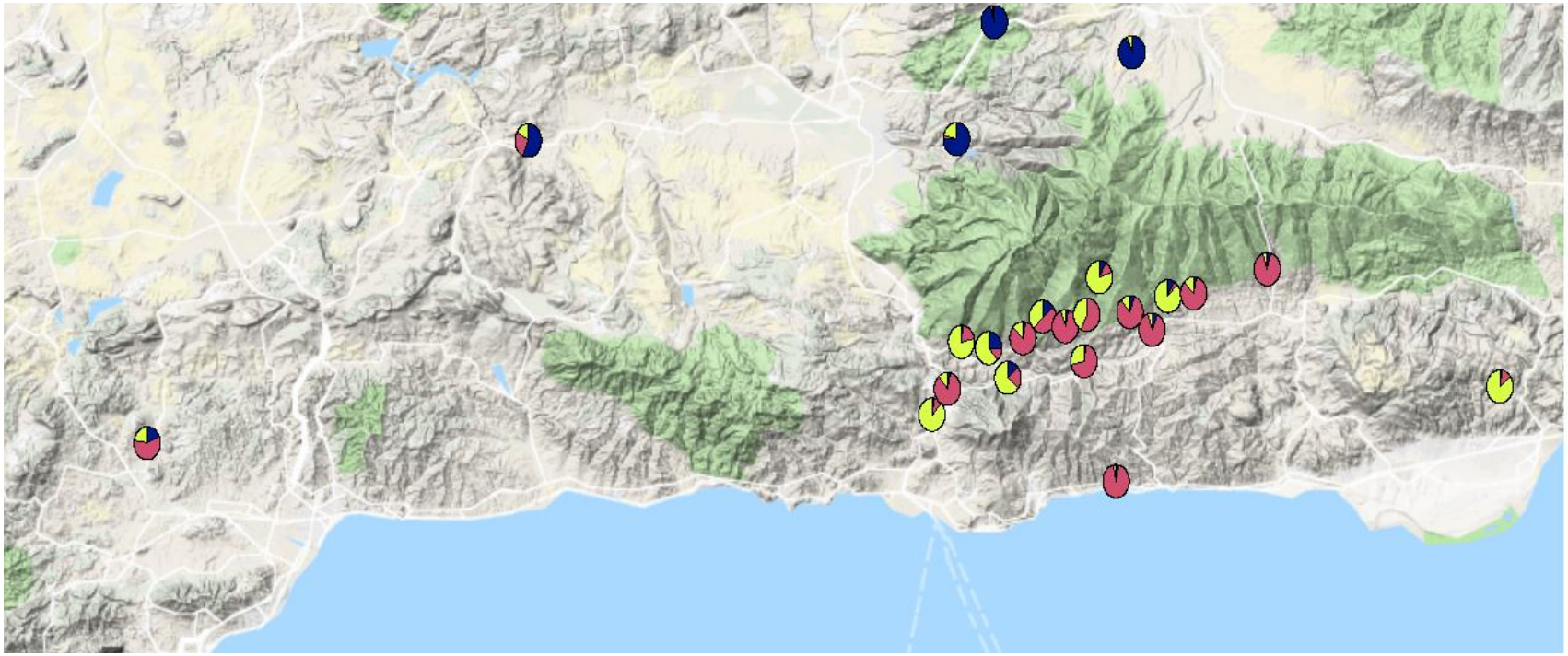


Figure 3.3. Geographic map showing patterns of genetic structure and admixture between populations using the Q values estimated with STRUCTURE at K=3. In the pie charts, blue colour represents the northern clade genotype, yellow colour represents the *Kickxiella* genotype from the southern clade and the purple colour represents the *Antirrhinum* genotype also from the southern clade.

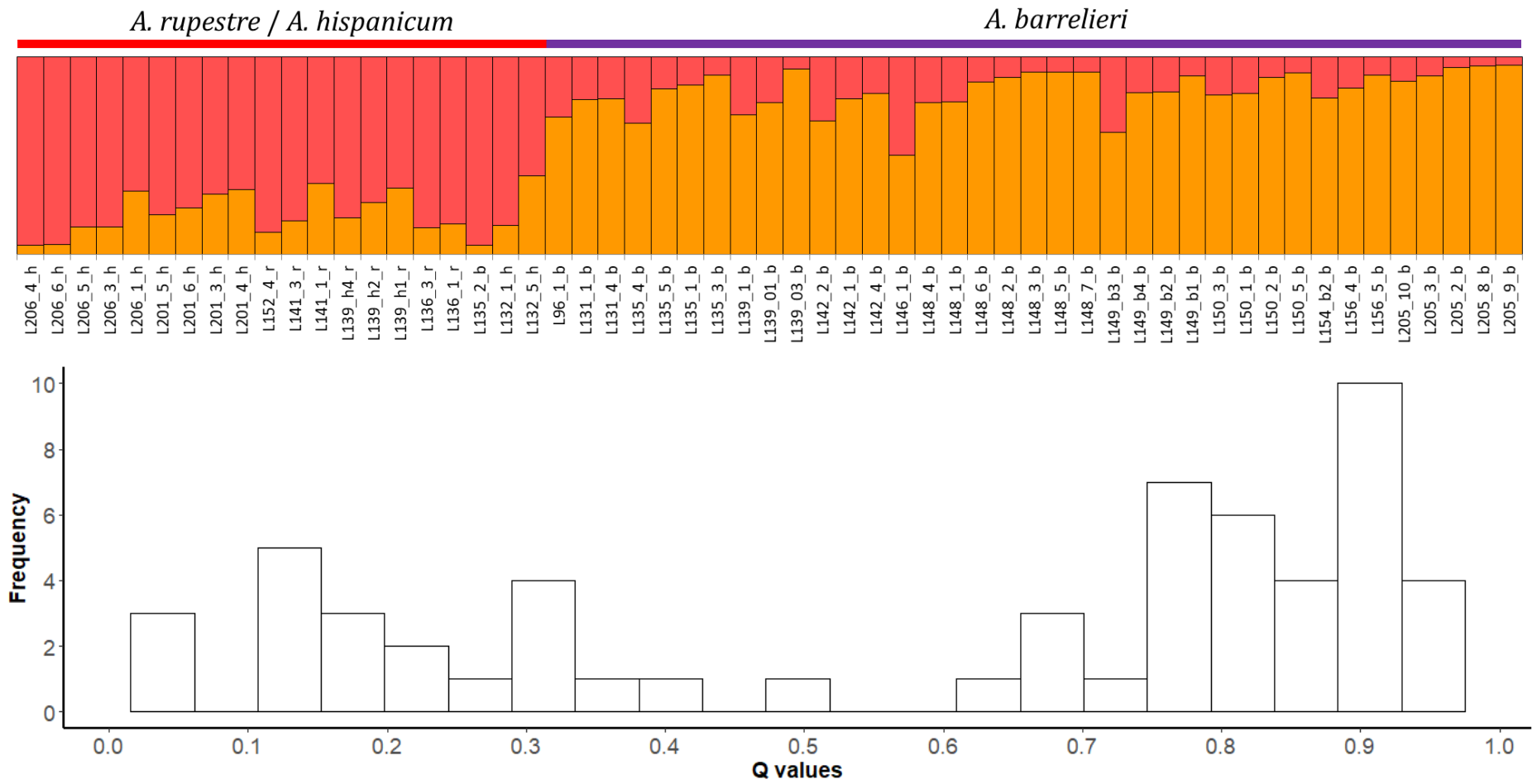


Figure 3.4. On top, bar plot showing the results from the hybridisation analysis in STRUCTURE at K=2 for the individuals from the Sierra Nevada. On the bottom, histogram of the distribution of Q values among the 57 individuals sampled as estimated by STRUCTURE.

Discussion

In this study I found a high proportion of hybrid individuals between the divergent species *A. barrelieri* and *A. rupestre* across populations in the south of the Sierra Nevada. Despite this high proportion of hybrid individuals found, intermediate phenotypes are uncommon in these populations. I suggest selection on phenotypic traits and ecological preferences could be maintaining morphological divergence between species despite hybridisation. Additionally, I characterized patterns of genetic structure in wild populations of *Antirrhinum* by analysing thousands of nuclear SNP loci. My results show a clear geographic structure revealing the Sierra Nevada acts as a barrier to gene flow between northern and southern populations.

Hybridisation

The results show the presence of abundant hybrid individuals between the morphological divergent species *A. barrelieri* and *A. rupestre* in the south of the Sierra Nevada. The proportion of hybrid individuals (52 from 57) represents an unexpected result as intermediate morphologies are uncommon in the wild (Wilson and Hudson, 2011). Nonetheless, my results show that the contact zone in the Sierra Nevada is actually composed of individuals with different degrees of admixed genetic ancestry and that individuals with intermediate Q values are scarce. Therefore it seems plausible that the establishment of new F1 hybrids may be rare and that individuals with admixed ancestry are either advance generation hybrids with phenotypes resembling one of the parental species or advance introgressants resulting from the backcross of a hybrid with one of the parental taxa. A similar case has been observed in a hybrid zone between the plant species *Iris fulva*, *I. brevicaula* and *I. hexagona*, in which different genotypic classes were abundant in the population but intermediate genotypes, particularly F1 hybrids, were absent. Hybrid individuals had phenotypes matching the parental species and were associated with specific habitats (Cruzan and Arnold, 1993). In *Antirrhinum*, Wilson *et al* (Unpublished) found an association between phenotypic classes and three ecological

variables in the Sierra Nevada. They found that individuals with *Kickxiella* phenotypes tend to live in habitats with a high slope, rocky substrates and low vegetation cover. In contrast, individuals with *Antirrhinum* phenotypes tend to live in deeper soils and lower slopes with a higher vegetation cover. This ecological preference suggests a role of selection in the maintenance of morphological divergence, with new hybrids being maladapted and failing to survive to reproduction.

My results prove a valuable comparison to a long running study of the Pyrenean hybrid zone between *A. striatum* and *A. pseudomajus*. These species have contrasting flower colours (*A. striatum* yellow, and *A. pseudomajus* magenta) with a steep cline in morphological and genetic variation. Evidence supports that differences are maintained by selection on pollinator preference as there are no clear differences in environmental conditions or pollinator distribution through the hybrid zone (Whibley *et al*., 2006; Bradley *et al*., 2017; Tavares *et al*., 2018). Although recombinants between flower colours have been found, F1 hybrids are rare (Tavares *et al*., 2018). Therefore it is likely that this hybrid frequency is reflecting an old history of contact in the hybrid zone and geographic limits for dispersal between pure parental taxa at the margins of the hybrid zone (David Field, personal communication). Unlike the hybrid zone between *A. straitum* and *A. pseudomajus*, which is composed of very similar species from the same subsection (and often considered subspecies of the same taxon), the species in my study are highly ecologically and morphologically divergent. As such, it is surprising that so many hybrids are formed.

Another possible explanation for the formation and establishment of hybrids in these populations is the abundance of human settlements in the south of the Sierra Nevada. In this region, sympatric populations of species *A. barrelieri* and *A. rupestre* are associated with anthropogenic disturbance like cities, roads, bridges and olive fields (personal observation). All these places have in common the presence of a wide set of environmental conditions suitable for both taxa. In the case of *A. rupestre*,

species from subsection *Kickxiella* are considered to be chasmophytes, with a need for surfaces with crevices to establish (Torres *et al*, 2013). This means that building walls and bridges within cities represent new available sites for this species to colonise. The distribution of *A. barrelieri* has been previously related to disturbed sites in the south of Spain such as walls, roads and roadside verges (Güemes, 2009). Therefore, it is possible that anthropogenic disturbance is increasing the frequency of hybridisation by the creation of habitats with intermediate ecological conditions and at the same time increasing the probability of establishment for new hybrids. Anthropogenic sites promoting hybridisation between previously ecologically isolated species has been documented before. For example, AFLP markers and morphological analyses showed that hybrid populations between three species of *Pericallis* were significantly more abundant in road verges in the Island of Tenerife (van Hengstum *et al*., 2012). In the United States and Canada, hybrid individuals between the species *Juglans cinerea* and *J. ailantifolia* were found growing only on roadsides, fencerows and streams with low levels of interspecific competition (Hoban *et al*., 2012).

In this work I was not able to identify loci underlying differentiation between *A. barrelieri* and *A. rupestre*. This limitation was caused by a significant amount of issues related to the sequencing approach. One of them was the high amount of missing data between samples that caused a reduction in sample size in several populations. A second issue, probably the most significant, was the unexpected low coverage obtained from sequencing (~2x). Both factors, missing data and low coverage, produced a pattern of low resolution in genome scans that was not adequate for the detection of important loci for divergence of populations and species. In future works a denser sampling of populations and a whole genome resequencing should be able to solve this limitations.

Patterns of genetic structure

My results show that geography has a strong influence on the genetic structure of *Antirrhinum* species in the south of Spain. Although previous studies have also detected a geographical correlation between genotypes and geography in *Antirrhinum*, these results have been based on the genetic relationships between chloroplast haplotypes and their incongruence with the morphological classification of the genus (Vargas *et al*., 2009; Wilson and Hudson, 2011). In this study, the amount of nuclear markers allowed us to visualize detailed patterns of genetic structure between populations on a wide geographic scale. One of these patterns is the clear division between populations distributed in the north and in the south of the Sierra Nevada, supporting the role of this mountain range as an important barrier to gene flow. Historically, mountain ranges in the south of Europe have played an important role shaping the current distribution of plants species. Different sources of evidence, like chloroplast markers and pollen maps, support the theory that during the last glaciation ice covered mountains were important barriers for the movement and colonisation of plants from different refugia in the south of Spain, Italy and the Balkans (Comes and Kadereit, 1998; Hewitt, 1999). This importance was also recognized by Vargas *et al*. (2009), who posits the mountains of the South East of Spain as a main centre of *Antirrhinum* speciation. Additionally, Rothmaller (1956) and Webb (1971) proposed that this geographic speciation was followed by multiple secondary contact events driven by climatic fluctuations in the Quaternary with chorological, morphological and ecological data supporting this scenario (Vargas *et al*., 2009). Hence it is very likely that the patterns of subpopulation structure found in this work do not reflect morphological variation between species but the evolutionary history of speciation in isolation followed by a recent event of secondary contact.

Other factors that could be affecting the genetic structure of *Antirrhinum* populations are biotic and abiotic conditions. The distribution of pollinators, for example, will have a big impact in *Antirrhinum* reproduction, as most of the species in the genus have a predominant multiallelic self-incompatibility system which prevents self-

fertilisation. The combination of this mating system and the morphology of the flower hiding the anthers from the exterior, makes pollinators essential for *Antirrhinum* reproduction and the maintenance of the genetic diversity of natural populations (Carrió and Güemes, 2014; Vargas *et al*., 2017). A good example of the importance of pollinator behaviour for the reproduction of *Antirrhinum* is the species *A. microphyllum*, distributed in the centre of Spain. Natural populations of *A. microphyllum* have been reported as having a short distance genetic structuring and has been related to the behaviour of its main pollinator. Female individuals of the territorial bee *Rhodanthidium sticticum* collect pollen from patches within 10m distance. Hence it is likely that some of the geographic variation in genetic structure in *Antirrhinum* will be explained by pollinator distribution and behaviour.

Seed dispersal is another important factor that can affect the genetic structure in *Antirrhinum*. A short range of dispersal combined with a limited distribution of suitable conditions for establishment can have a massive influence in the amount of genetic diversity in plant populations. In *Antirrhinum,* seed dispersal occurs by a combination of gravity and wind, with seeds not falling far from the parental plants (Torres *et al*., 2003; Carrió *et al*., 2009). This method of dispersal is considered to promote differentiation between populations by increasing the likelihood of inbreeding within populations (Loveless and Hamrick, 1984). However, in *Antirrhinum* this factor will have a limited effect on genetic diversity as the strong self-incompatibility system in most taxa will reduce the inbreeding in great extent. On the other hand, the presence of suitable environmental conditions for germination and establishment is more likely to influence the observed genetic structure. Wilson and Hudson (Unpublished) report a strong a relationship between the morphology of the main subsections and the environmental variables slope, substrate type and vegetation cover. Plants with *Kickxiella* morphologies are found on sites with rocky surfaces, high slopes and low vegetation cover while plants with *Antirrhinum* morphologies prefer level soils with higher vegetation cover. In the species *A. microphyllum* (subsection *Kickxiella*), Torres *et al*. (2003) argues that the

genetic structure can be related to the abundance and distribution of cliffs and crevices necessary for the anchorage of the roots. Therefore an uneven distribution of suitable environmental conditions across the landscape in combination with a limited dispersal of seeds could be the cause of the patterns of genetic structure found in this work.

Conclusion

I found high levels of hybridisation between the species *A. barrelieri* and *A. rupestre* in populations from the south of the Sierra Nevada. Despite high levels of hybrid ancestry, individuals with intermediate morphologies are rare. Habitat preferences could be responsible for the maintenance of morphological divergence between species. However, anthropogenic disturbance and the creation of new environments could also be responsible for the establishment of hybrid populations in the wild. The genetic structure of *Antirrhinum* populations in the south of Spain showed that the Sierra Nevada plays an important role in the evolution of populations. The mountains could be acting as barriers to seed dispersal or pollinator movement.

Chapter 4

Genomics of adaptation in *Antirrhinum* inferred from QTL mapping.

Genomics of adaptation in *Antirrhinum* inferred from QTL mapping.

Abstract

Understanding the genetic mechanisms of adaptation is fundamental to understand speciation. Factors like the number, distribution and phenotypic effects of loci controlling adaptive traits can promote or hinder the diversification process. The main goal of this chapter was to identify the genetic architecture of adaptive traits between the hybridising species *A. barrelieri* (subsection *Antirrhinum*) and *A. rupestre* (subsection *Kickxiella*) with the aim of understanding the genetic mechanisms keeping the morphological divergence between species. First, I performed a QTL analysis in an F2 cross derived from individuals from the hybrid populations in the Sierra Nevada using restriction site associated DNA markers. I found most of the traits are controlled by multiple regions across the genome with medium and small effects suggesting a role of natural selection in the divergence of the species. Second, I compared my results with data from a previous cross between the species *A. charidemi* and *A. majus*. These species are not closely related and do not have overlapping distributions, hence are expected to have evolved in allopatry. This comparison showed fewer QTL involved in each trait in the species from the Sierra Nevada, likely caused by the homogenizing effects of gene flow. The identification of shared regions bearing multiple QTL between both studies suggest these sites are important for the divergence of the morphological subsections.

Introduction

Exploring the genetics of architecture of adaptation is fundamental to understand the process of speciation and evolution. Adaptive divergence of populations and species can arise via a number of genetic mechanisms. Firstly, adaptation can involve genetic changes at a number of loci spread through the genome. Secondly, genetic changes at a particular locus or loci in close linkage (e.g. multi-gene complexes or 'supergenes') may underlie adaptation. While a number of high profile cases exist where speciation occurs with few genetic changes, such as in high Müllerian mimic *Heliconius* butterflies (Booker *et al*., 2015), these may be a special case, and it may be more likely that divergence involves regions across the genome. Using genomic approaches allow us to test between competing models of divergence and to characterise the landscape of speciation and adaptation.

A major breakthrough in the study of the genetics of adaptation was the development of genome wide association methods like QTL analysis. A QTL or quantitative trait locus is a genomic region containing the alleles responsible for the variation in a quantitative trait. A good example of quantitative traits are height and weight (Collard et al., 2005). The aim of a QTL analysis is the detection of these regions and consist in the association between phenotypic and genotypic variation in an artificial cross (Broman and Sen, 2009). This method, when applied to natural populations, can be used to locate the genomic regions responsible for phenotypic differences between populations. Despite this theoretical framework it is only relatively recently that it has been common place to test these models with new experimental data. The development of new experimental approaches like quantitative trait loci (QTL) analysis and the recent genomic revolution have allowed the collection of unprecedented amounts of data about the genetics of speciation. In a QTL analysis it is possible to estimate the number, position, effect sizes, directions and epistatic interactions of genomic regions responsible for important morphological traits. In addition, we can combine these data with ecological,

demographic and phylogenetic analyses to make inferences about the main factors affecting adaptation and speciation in natural populations.

A main issue in identifying the genetics of adaptation it is that it relies on the necessity of an organisms that are ecologically tractable. Hence most studies have focused on a small set of model organisms like *Heliconius* butterflies (Huber *et al*., 2015), stickleback fish (Marques *et al*., 2016) and plants like *Arabidopsis thaliana* (Price *et al*., 2018) and *Mimulus guttatus* (Ferris *et al*., 2017). It is evident that we need to address questions related to the genetic bases of ecological divergence in a broader set of organisms if we ever want to understand the general mechanisms underlying speciation.

The genus *Antirrhinum* is a classical model system for plant development and it is also emerging as a promising system for studying population dynamics and natural hybridization. The genus is composed of around 20 species and almost all are capable of hybridizing and producing fertile offspring. Traditionally the genus has been divided into different morphological subsections (Rothmaler, 1956). Species in subsection *Kickxiella* are short with a prostrate habit, hairy stem, ovate leaves and white axillary flowers while species in subsection *Antirrhinum* are tall erect plants with lanceolate leaves and no hairs outside the terminal inflorescence of large magenta flowers. The morphological differences between subsections are likely driven by ecology, with plants from subsection *Kickxiella* growing in siliceous soils on cliffs and wall surfaces with low vegetation cover, and plants from subsection *Antirrhinum* grow in deeper calcareous soils with a higher vegetation cover (Wilson et al., 2016).

Previous work has examined the genetic bases of variation in an experimental cross between *A. majus* (subsection *Antirrhinum*) and *A. charidemi* (subsection *Kickxiella*), with a focus on allometry of leaves and flowers (Feng *et al*., 2009). These species are unrelated (chapter 2) and do not have overlapping distributions, hence is expected that the morphological divergence between these taxa happened in isolation. Here I build on this previous work by characterising QTL responsible for trait differences

between two closely related species of snapdragons (chapter 2). The species *A. rupestre* (subsection *Kickxiella*) and *A. barrelieri* (subsection *Antirrhinum*) have typical morphology for each respective subsection but are characterised by low genetic differentiation (Wilson and Hudson, 2011), and are found to recurrently hybridise in the field (Chapter 3). High morphological divergence in the face of low genetic differentiation may be explained by natural selection via habitat preferences, with individuals with intermediate morphologies being selected against in natural populations. The comparison of the genetic architecture of adaptive traits between studies represents an excellent opportunity to explore the influence of gene flow in the evolution of the genetic architecture of adaptation between species diverging in sympatry. Theory predicts that different evolutionary forces will be acting in populations in allopatry causing a more heterogeneous pattern of divergence that the one expected in hybridizing populations in sympatry (Nosil *et al*., 2009). Therefore although I expect to find some of the QTL reported by Feng *et al*. (2009), I also expect to detect different sets of loci associated with each character. Here I used restriction site associated DNA markers to perform a QTL analysis aiming to identify the number of QTL involved in each trait, their location and effect sizes. Then I infer patterns of evolution related to divergence in sympatry by comparing my results with a cross between species distantly related (Feng *et al*., 2009).

METHODS

Study species

The species *A. barrelieri* (subsection *Antirrhinum*) and *A. rupestre* (subsection *Kickxiella*) were chosen for mapping parents as a previous genetic study reports a low genetic distance between them and a high amount of shared bands at AFLP markers (Wilson and Hudson, 2011; Wilson *et al*., 2016, chapter 4). Each species has a typical morphology of subsection *Antirrhinum* and *Kickxiella* respectively and therefore they differ in several important taxonomic traits. This means they provide

a direct comparison to the previous mapping cross between *A. majus* and *A. charidemi* by Feng *et al*., (2009). *A. barrelieri* has an upright habit with a glabrous stem, lanceolate leaves and a terminal inflorescence with magenta flowers. In contrast, *A. rupestre* is characterised by prostrate habit with stem and leaves covered by trichomes, ovate leaves and white axillary flowers (figure 4.1). In a previous study by Wilson *et al*. (Unpublished), a relationship between morphology and certain environmental variables was found. Individuals with a *Kickxiella* like morphology are found in cliffs and rocky surfaces with low vegetation cover. In contrast, individuals with *Antirrhinum* morphology prefer sites more level soils with higher vegetation cover.

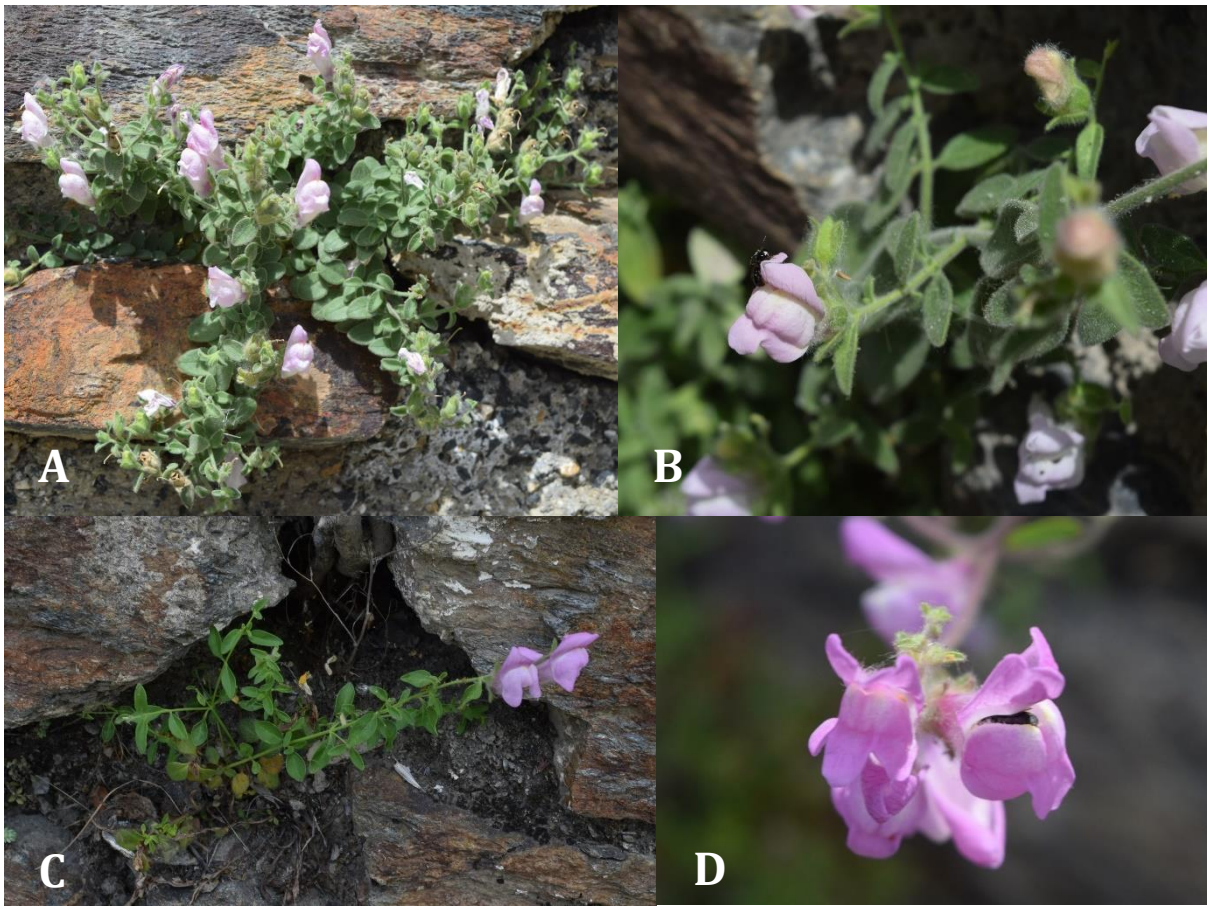


Figure 4.1. (A) Growth habit and morphology of *A. rupestre* a species with the typical *Kickxiella* morphology. (B) Close caption of *A. rupestre* flower. (C) The species *A. barrelieri* with a typical *Antirrhinum* morphology. (D) Close caption of *A. barrelieri* flowers.

Plant material

A single family F2 population of 230 individuals was produced by Andrew Hudson using wild collected accessions of *A. barrelieri* and *A. rupestre* as the parents (see complementary material for accession locations). The accessions used were chosen based on the phenotypes, selecting the samples that had the biggest difference in measurements in most of the traits of interest (e.g. tall plants vs short plants, long leaves vs short leaves, many flowers vs few flowers, etc.).

An individual of *A. rupestre* was crossed with an individual of *A. barrelieri* reciprocally to produce an F1 generation. Given that both species are self-incompatible, the F2 was generated by crossing two compatible F1 individuals reciprocally. All plants were grown in the glass-houses at the Institute of Molecular plants Sciences, University of Edinburgh following the growth conditions specified in Hudson *et al*., (2008b).

Data Capture and Measurements

Twelve vegetative and reproductive traits were chosen based on the criteria described in Wilson and Hudson (2011). Phenotypic traits and their abbreviations are listed in table 1. All plants were measured after the first five flowers on the inflorescence were open. For leaf characters, a single leaf from each main stem internode was collected including the petiole. Leaves were arranged and pasted on a grid in an A4 piece of paper and digitalized using a Nikon Coolpix 4500 digital camera. The leaf size and shape were analysed digitally. Flowers were collected from each plant including the pedicel and then digitalized. Flowers were then dissected separating pedicel, sepals, petals, stamens and pistil. The corolla was split into dorsal and ventral faces and then digitalized after pasting on a grid in the same way as leaves. Measurements from each part of the flower were taken from the digital images. All measurements were made to nearest 0.01 mm.

Table 4.1. Phenotypic characters used in the QTL analysis.

Phenotypic characters	Description
IntLen	Cumulative value of the length of ten internodes in cm.
BrLen	Length in cm from the first node to the first flower in a lateral branch.
BrInd	Ratio between the number of nodes until the first flower in the main stem and a lateral branch.
LeafLW	Average ratio between the length and width of the leaves measured on 10 leaves in cm.
LeafA	Average area of 10 leaves in mm^2.
NumFlowers	Number of floral buds produced.
PedLen	Pedicel length in cm.
FloLen	Length of the flower in cm.
SepL	Sepal length in mm.
SepW	Sepal width in mm.
DpAngle	External angle between the petal lobe and the corolla tube.
DpInner	Internal angle between the petal lobe and the corolla tube.

Genotypic data collection

DNA extraction and library preparation were performed as specified in chapter 2, but using *Sbf I* instead of *Pst I* as the restriction enzyme. Libraries were sequenced at Edinburgh Genomics on a HiSeq-4000 generating 176,770,163 paired end sequencing reads.

Construction of the genetic map

Raw reads were filtered and aligned to the *A. majus* genome using the alignment software GSNAP (Wu and Nacu, 2010) and the core components of the software STACKS (Catchen *et al*., 2013) by Edinburgh Genomics in 2013. Because the version

2 of the *A. majus* genome was only available last year, these early stages of analysis were done using the *A. majus* genome version 1.

After the removal of individuals with high amounts of missing data (>80 %), loci within 1 kb intervals were grouped into multilocus markers to represent single different *Sbf* I sites. As a result, the recombination map was built with a total of 190 individuals and 710 multilocus markers using the software JoinMap and a maximum likelihood algorithm with the default settings. This part of the analysis was done by Matthew Barnbrook as part of his dissertation project. The resulting recombination map was compared to the sequence of the version 2 *A. majus* genome (LI *et al*., 2018) to identify the orientation of each linkage group relative to the reference genome.

QTL analysis

Quantitative trait loci were detected using R/qtl (Broman *et al*., 2003) using standard interval mapping and Haley-Knott regression for two and multiple QTL analyses (Lander and Botstein, 1989). Because all the species within *Antirrhinum* are diploids and self-incompatible each marker can have up to four alleles. However R/qtl is designed for inbreeding organisms with only two alleles per marker. Therefore I filtered out markers with three or more alleles before the analysis leaving a total of 261 multi locus markers. I followed the workflow specified in Broman and Sen (2009) starting with a genome scan with a single QTL model in order to identify loci with large effects and then a two dimensional scan searching for interactions or loci with small marginal effects. In both scans I evaluated the statistical significance of each QTL using a permutation test with 1000 replicates. These results were used as starting point in a multiple-QTL analysis using a stepwise model selection approach with a penalty of 5% significance threshold. For each QTL, I estimated the individual and total percentage of variance explained (PVE), an approximate 95% Bayesian credible interval, and the additive and dominant effects. The results were visualized in the recombination map using the R package LinkageMapView (Ouellette *et al*., 2018).

Comparison with previous results

The results obtained from the QTL analysis were compared with the results shown in Feng *et al* (2009) for five morphological traits. Although their published work is mainly focused on the evolution of allometry in leaves and flowers, I was provided unpublished data by A. Hudson with a wider range of morphological traits. Here, I compared the genomic architecture of the traits internode length, leaf length and width, number of flowers, pedicel length, and flower length. In order to compare the results of the QTL analysis with the results found by Feng *et al* (2009), I compared the recombination maps from both studies to the *A. majus* reference genome version 2 (Li *et al*., 2019). This comparison was also done in order to determine the order and orientation of the linkage groups in both recombination maps.

RESULTS

Recombination map

A total of 261 markers were mapped into eight different linkage groups (LG) with a total map length of 593 cM (figure 4.2). The number of linkage groups formed agrees with the number of chromosomes in *Antirrhinum*. I found an uneven distribution of markers between linkage groups having LG1 the highest amount of markers (57) and LG7 the fewest (19).

QTL mapping

Number and location

A total of 21 significant QTL were found for eight of the twelve morphological traits. My results show that most traits are controlled by more than one QTL (figure 4.1). The average length and width of leaves are traits with the highest number of QTL with six regions detected (table 4.2). In contrast, I only detected a single QTL responsible for internode length, and branch length. Two regions in LG3 and LG4 were found having more than two overlapping QTL for different traits. The QTL observed in these regions are for the length and size of flowers and leaves.

Percentage of Variance Explained and epistatic interactions

I found three QTL with large effects (>20 PVE), eight with intermediate effects (>10 PVE) and 9 with small effects (<10 PVE). The QTL with the biggest effect observed was detected in LG8 for the trait flower length, with a single QTL explaining 34.4% of the variation. I identified QTL responsible for more than 50 PVE of the traits average length and width of leaves, and flower length.

From the 21 quantitative trait loci detected, I found four epistatic interactions between eight different sites. The rest of the QTL detected have additive effects. The epistasis between LG1 and LG7 is especially important for the trait sepal width, as there was not a significant effect of each individual QTL without including the interaction.

Direction and dominance

The QTL directions can be predicted by the phenotypes of the parental species in all of the traits. For example, *A. barrelieri* alleles produce individuals with longer branches and bigger and more numerous flowers, while *A. rupestre* alleles produce shorter individuals with smaller and less numerous flowers. Both species contribute dominant alleles in a similar number of loci with *A. rupestre* being dominant in 12 loci and *A. barrelieri* being dominant in 9. I also found six QTL with overdominant effects. In other words, individuals with heterozygous genotypes having higher phenotypic values than both homozygotes.

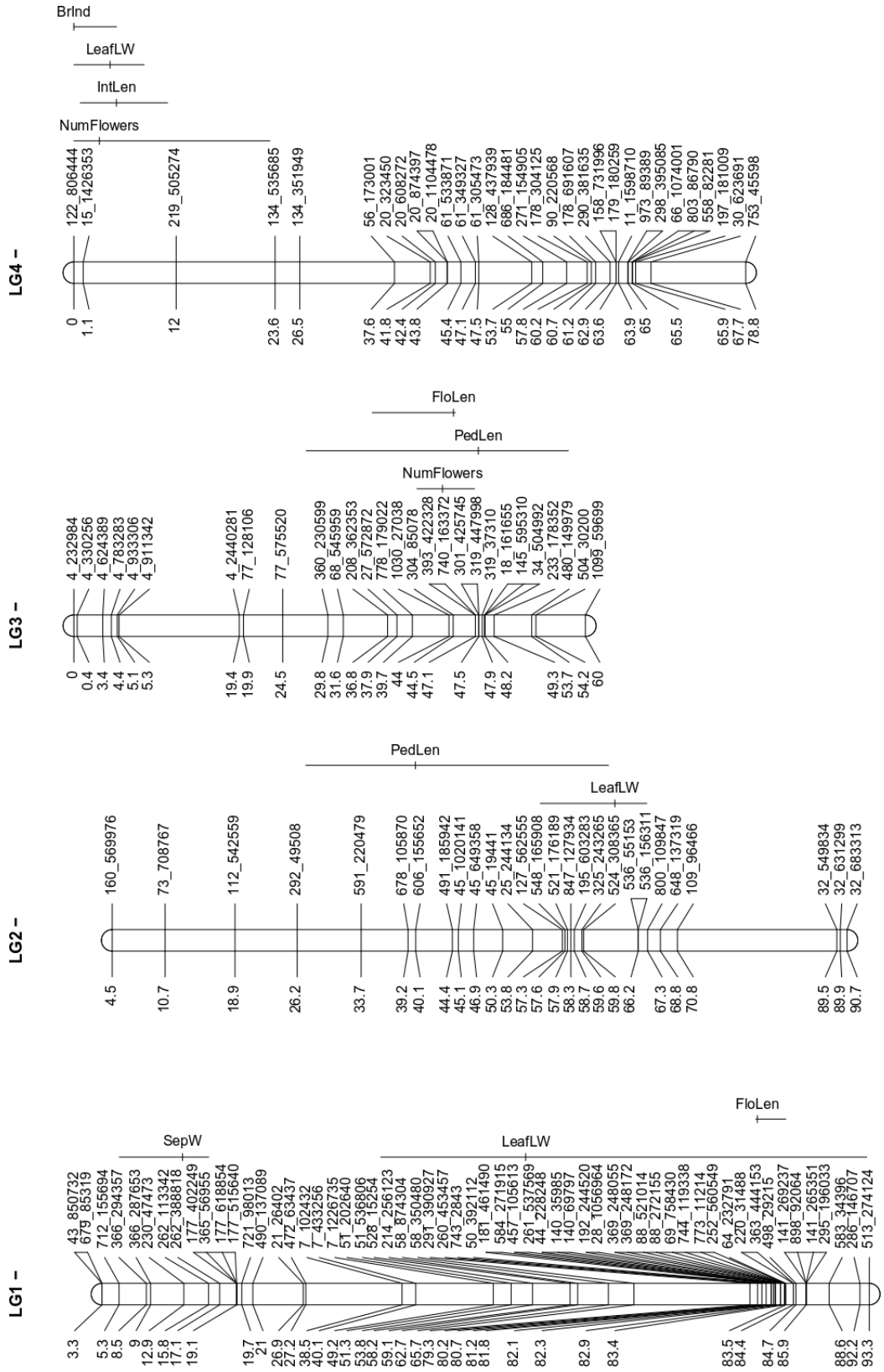
Comparison with previous QTL

I found 12 QTL on similar positions for the five characters used compared between my *A. barrelieri* x *A. rupestre* cross and the previous *A. charidemi* x *A. majus* cross (table 4.3). All the QTL were found on the same linkage groups although in variable positions. In all of these traits fewer QTL were involved in comparison to Feng *et al*. (2009), but they had bigger individual effect sizes and explained similar amounts of total PVE per trait.

I also found support for the existence of two important sites in LG3 and LG4 with multiple overlapping QTL in both studies. The first one is involved in floral traits like size and number of flowers. The second is related to architectural characters like plant height, shape and size of the leaves and the branching pattern.

Table 4.2. Summary of the significant QTL found for all traits using 261 multilocus RAD markers in Rqtl (Broman *et al.*, 2003). Letters on top of LG numbers represent interactions, with matching letters indicating the loci interacting.

Trait	Total PVE	LG	Position	Interval (cM)	LOD	Individual PVE	Additive effect	Dominance Effect
IntLen	17.52	4	5	0.75 - 11	7.74	17.52	-2.24	2.6
BrLen	12.73	6	29.7	17.49 - 41.4	5.53	12.73	12.64	-0.75
BrInd	24.56	4	0	0 - 5	7.49	16.68	-0.55	0.1
BrInd		8	47.33	37.58 - 55.27	3.43	6.8	-0.17	0.46
LeafLW	63.36	1	86	36 - 93	5.41	6.94	0.22	0.04
LeafLW		2[a]	57.5	54.71 - 67.21	5.3	6.73	0.2	0.06
LeafLW		4[b]	0	0 - 8.25	3.9	3.99	0.13	0.06
LeafLW		5	36.6	9.56 - 36.31	2.98	3.68	0.14	0.06
LeafLW		6[a]	36.7	22.65 - 31.15	9.16	12.23	0.27	-0.14
LeafLW		7	35.2	23.19 - 48.19	6.49	8.37	-0.21	-0.06
LeafLW		8[b]	51.8	51.08 - 54.33	12.22	16.97	0.19	-0.38
NumFlowers	25.74	3	47	40.25 - 47	7.49	19.37	1.38	0.71
NumFlowers		4	12	0 - 23	3.24	7.84	0.87	-0.76
PedLen	32.4	2	40.07	27.21 - 62.71	5.55	14.32	-0.1	0.03
PedLen		3	47.45	27.25 - 58	2.69	6.59	-0.05	0.04
PedLen		6	61.65	57.4 - 66.81	3.52	8.78	0.08	-0.04
FloLen	51.41	1	80.17	80 - 83.50	14.66	31.82	0.33	0.27
FloLen		3[b]	44.75	35 - 45.75	8.36	16.17	0.09	0.1
FloLen		7[b]	30.69	29.19 - 34.69	15.58	34.4	-0.06	-0.3
SepW	26.1	1[a]	12.9	5.33 - 15.82	8.63	23.46	-0.09	0.18
SepW		7[b]	65	51.21 - 65	8.69	23.64	0.05	0.16


LG4
BrInd
LeafLW
IntLen
NumFlowers
122_806444
15_1426353
219_505274
134_535685
134_351949
56_173001
20_323450
20_608272
20_874397
20_1104478
61_533871
61_349327
61_305473
128_437939
686_184481
271_154905
178_304125
90_220568
178_691607
290_381635
158_731996
179_180259
11_1598710
973_89389
298_395085
66_1074001
803_86790
558_82281
197_181009
30_623691
753_45598
0
1.1
12
23.6
26.5
37.6
41.8
42.4
43.8
45.4
47.1
47.5
53.7
55
57.8
60.2
60.7
61.2
62.9
63.6
63.9
65
65.5
65.9
67.7
78.8
LG3
FloLen
PedLen
NumFlowers
4_232984
4_330256
4_624389
4_783283
4_933306
4_911342
4_2440281
77_128106
77_575520
360_230599
68_545959
208_362353
27_572872
778_179022
1030_27038
304_85078
393_422328
740_163372
301_425745
319_447998
319_37310
18_161655
145_595310
34_504992
233_178352
480_149979
504_30200
1099_59699
0
0.4
3.4
4.4
5.1
5.3
19.4
19.9
24.5
29.8
31.6
36.8
37.9
38.7
44
44.5
47.1
47.5
47.9
48.2
49.3
53.7
54.2
60
LG2
PedLen
LeafLW
160_569976
73_708767
112_542559
292_49508
591_220479
678_105870
606_155652
491_185942
45_1020141
45_649358
45_19441
25_244134
127_562555
548_165908
521_176189
847_127934
195_603283
325_243265
524_308365
536_55153
536_156311
800_109847
648_137319
109_96466
32_549834
32_631299
32_683313
4.5
10.7
18.9
26.2
33.7
39.2
40.1
44.4
45.1
46.9
50.3
53.8
57.3
57.6
57.9
58.3
58.7
59.6
59.8
66.2
67.3
68.8
70.8
89.5
89.9
90.7
LG1
FloLen
SepW
LeafLW
43_850732
679_85319
712_155694
366_294357
366_287653
230_47473
262_113342
262_388818
177_402249
365_56955
177_618854
177_515640
721_98013
490_137089
21_26402
472_63437
7_102432
7_433256
7_1226735
51_202640
51_536806
528_15254
214_256123
58_874304
58_350480
291_390927
260_453457
743_2843
50_392112
181_461490
584_271915
457_105613
261_537569
44_228248
140_35985
140_69797
192_244520
28_1056964
369_248055
369_248172
88_521014
88_272155
69_758430
744_119338
773_11214
252_560549
64_232791
220_31488
363_444153
498_29215
141_269237
898_92064
141_265351
295_196033
583_34396
286_146707
513_274124
3.3
5.3
8.5
9
12.9
15.8
17.1
19.1
19.7
21
26.9
27.2
38.5
40.1
49.2
51.3
53.8
58.2
59.1
62.7
65.7
79.3
80.2
80.7
81.2
81.8
82.1
82.3
82.9
83.4
83.5
84.4
84.7
85.9
88.6
92.2
93.3

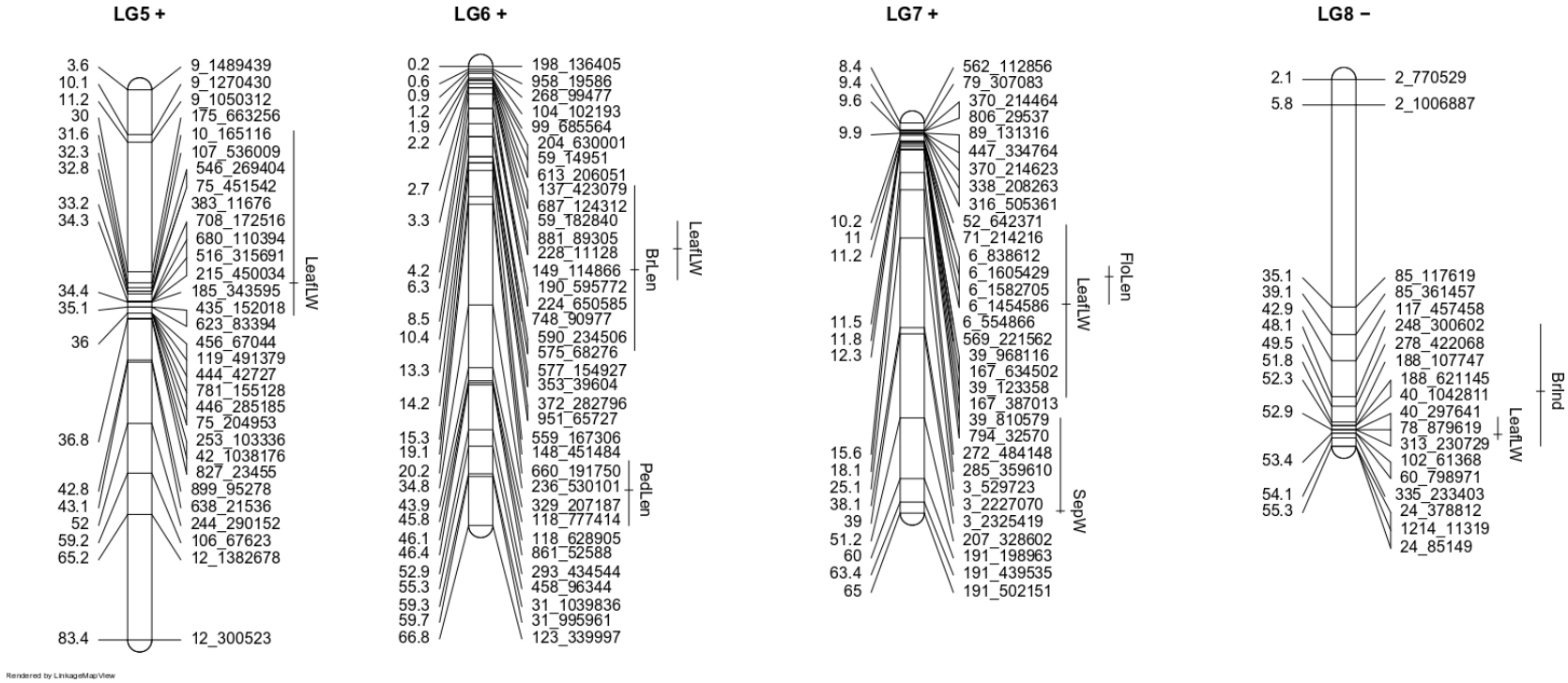


Figure 4.2. Recombination map produced with 261 RAD markers showing the location of QTL found. The name of the LG correspond to the chromosome number in the reference genome. The symbol next to the name correspond to the orientation, with a plus sign meaning that the position of the markers matches their position in the reference genome and a minus sign means they are in opposite direction.

Table 4.3. Matching results between this work and Feng *et al.* (2009). The names of the traits measured, the linkage groups, the estimated location of each QTL, the percentage of variation explained and the orientation of each LG are shown on the table. Plus and minus signs represent the orientation of the LG with respect of the reference genome. A plus sign means that the LG has the same orientation as the reference and a minus means the LG is in opposite direction. A result was considered a match only when a QTL responsible for a similar trait was located in the same linkage group and in a similar relative position.

Current work					**Feng *et al.* (2009)**				
Trait	**LG**	**Position**	**PVE**	**Orientation**	**Trait**	**LG**	**Position**	**PVE**	**Orientation**
IntLen	4	5	17.52	-	Plant height	4	87	2.80	+
LeafLW	1	86	6.94	-	Leaf area	1	20	2.26	+
LeafLW	5	36.6	3.68	+	Leaf area	5	50	18.67	+
LeafLW	4	0	3.99	-	Leaf area	4	25	7.08	+
LeafLW	2	57.5	6.73	-	Leaf area	2	56	3.12	-
LeafLW	6	36.7	12.23	+	Leaf area	6	51	7.49	+
LeafLW	8	51.8	16.97	-	Leaf area	8	21	5.13	+
NumFlowers	4	12	7.84	-	Flowers on 31 August	4	43	6.82	+
NumFlowers	3	47	19.37	-	Flowers on 31 August	3	38	4.26	+
PedLen	6	61.65	8.78	+	Pedicel length	6	22	7.49	+
FloLen	1	80.17	31.82	-	Petal tube length	1	33	6.30	+
FloLen	7	30.69	34.4	+	Petal tube length	7	26	10.70	+

Discussion

In this chapter I identified genomic regions involved in the morphological divergence of the species *A. barrelieri* and *A. rupestre*. The results showed most traits are controlled by several genome wide distributed QTL with different effect sizes and found regions with multiple overlapping QTL controlling different traits. After comparing my results with a previous cross between a pair of unrelated species I found fewer QTL involved per trait in the populations from the Sierra Nevada. This is likely caused by homogenising effects of gene flow as the experimental variables between studies were very similar. The identification of matching regions bearing multiple QTL in both studies could suggest these sites are important for the divergence of the morphological subsections.

The genomics of adaptation

The results show that most of the morphological traits are controlled by several genome wide distributed QTL of intermediate and small effects (<20 PVE). This distribution and effect sizes have been hypothesised in taxa undergoing adaptation. Intermediate sized mutations are expected to be fixed first at the beginning of the adaptive process, as they are less likely to be lost by drift and more likely to contribute to adaptation. As the population moves closer to the fitness optimal, it is expected that the next mutation fixed will likely have a smaller effect size in comparison to the previous one. This means that mutations with small contributions to the overall phenotype will be more abundant when the population is next to the adaptive optimum (Kimura, 1983; Barton, 1998; Orr, 2005; Yeaman and Otto, 2011). The QTL with the biggest effect sizes (>20 PVE) were observed in traits involved in floral morphology. Recent theoretical work suggests that traits under selection for biotic factors will have QTL with relative higher effects than those under selection for abiotic factors, especially in populations experiencing gene flow. This is caused by the spatial and temporal variation in biological interactions leading to a more intricate adaptive landscape with several fitness peaks. In this scenario the fixation

of QTL of large effect will be favoured to facilitate the jump between different peaks (Yeaman and Whitlock, 2011). Results consistent with that theory were found in a meta-analysis of mapping studies in plants, where QTL controlling traits related to biotic interactions (e.g. pollination attraction, disease resistance) had bigger effects than traits involved in abiotic interactions [e.g. plant height, germination rates; (Louthan and Kay, 2011)]. This was also true for *Antirrhinum* mapping study by Feng *et al*. (2009), as they found the biggest effects in the QTL responsible for petal pigmentation (41 PVE) and petal cell area (21.43 PVE).

My results also suggest a role of developmental constrains in the evolution of different morphological forms. I found two regions in LG3 and LG4 with more than two overlapping QTL related to flower and leaf length and size. Although the method followed in this work did not allow the identification of QTL controlling more than one character, there is a wide body of evidence showing that traits related to organ size and shape in plants share developmental constrains (Mizukami, 2001; Ingram and Waites, 2006; Anastasiou and Lenhard, 2007). For example, in *Antirrhinum*, it has been observed that the class II TCP protein CINCINNATA has opposite effects on growth in leaves and petal lobes (Crawford, 2004). These two regions were also found by Feng *et al* (2009), supporting an important role for evolutionary constraint in the morphological divergence in *Antirrhinum* and especially in the hybrid populations.

The comparison between my results and the analysis done by Feng *et al* (2009), showed 12 QTL on similar positions for five traits. On these traits a reduction in the number of QTL involved can be observed despite explaining similar amounts of total variation per character. This results could be explained by differences in the experimental design. However, this is very unlikely as a higher number of markers (261 vs174), the same kind of cross (F2) and a similar amount of F2 individuals (190 vs 175) were used in both QTL analysis. Another explanation is that this reduction in the number of QTL observed is result of the homogenising effects of gene flow between species. Previous works using computer simulations, have found this

scenario with models of adaptation under selection-migration balance. For example, Griswold (2006) predicted that alleles responsible for phenotypic differences will have greater effects with migration between populations, as they segregate on average more frequently between populations than alleles with small effects. Yeaman and Otto (2011) proposed that after a few generation of adaptation with migration, small effect alleles could be replaced with fewer large effect alleles, hence explaining the reduction found in the number of QTL per trait.

Another way to explain the differences in the number of QTL detected between this work and Feng *et al* (2009) is by some limitations in the QTL method. First, in this work we used two outbred specimens grew from seeds collected from natural populations. In contrast, Feng *et al* (2009) used an inbred cultivar of *A. majus* and an outbred *A. charidemi* individual grew from seeds collected in the wild. In a QTL analysis performed with inbred lines, minor allele frequencies (MAF) tend to be close to 0.5. However, the use of outbred species implies that some alleles can have much lower allele frequencies, therefore reducing the power of the analysis to detect small effect QTL's (Slate, 2013). A second limitation in the QTL analysis that could affect the results found in this work is the so called Beavis effect (Beavis and Beavis, 1994). This term refers to the phenomenon in which QTL analyses with small sample sizes (<500 individuals) not only have a reduced power of detection of small effect QTL's but also tend to overestimate the effect of QTL with statistical significance. This could mean that traits with no QTL detected in this work are actually polygenic, with several small effect QTL's being responsible of the differences between phenotypes. Together, the Beavis effect and the use of an outbred cross for the QTL analyses, could explain the big difference in the amount of small QTL's found between this work and Feng *et al* (2009). Nonetheless, despite the implications of these limitations in the number and effect sizes of the QTL found in this study, they do not imply that the QTL detected are not real. QTL analyses are only the first steps in the detection of candidate genes for important morphological traits and therefore more work is required in order to confirm the existence of these genomic regions.

Finally, my results could also be discussed from the perspective of an adaptive walk in a fitness landscape. In a shifting environment, it has been suggested that populations will spend more time away from the fitness optima. In consequence, loci acting in different directions and QTL will large effects will be more common (Louthan and Kay, 2011). In this work, I propose a similar scenario occurring in the populations in the Sierra Nevada, where the wide variety of environmental conditions created by the divergent habitats the site could lead different populations to different local fitness peaks. The two set of phenotypic traits in *A. barrelieri* and *A. rupestre* would represent two peaks on the fitness landscape. Gene flow between populations could allow the colonization of new habitats with different adaptive optima. Large-effect QTL would allow the transition between different adaptive peaks without passing through a low fitness valley and small-effect sized QTL would play a fundamental role in the fine tuning of phenotypic traits to these new conditions. A similar hypothesis has been proposed by Whibley *et al* (2006) for alleles controlling flower colour in populations from the Pyrenees of *A. pseudomajus* and *A. striatum*. In this hybrid zone between plants with yellow and magenta flowers, divergent selection was suggested as the main force maintaining a step cline in allele frequencies between taxa.

Conclusion

The analysis of the genetic architecture of *Antirrhinum* shows important regions for the divergence of species in isolation and in sympatry. These regions are responsible for traits related to the length and size of flowers and leaves and could be important candidates for outlier loci in studies of divergence. I also found a reduction in the number of QTL per trait but and an increase in the individual effect size in comparison to a previous work done on species evolving in isolation. This suggests that gene flow has an important role of in the evolution of the genetic architecture of adaptation in *Antirrhinum* and support a role of selection in the maintenance of morphological divergence in the hybrid populations.

Chapter 5
General discussion

Genomic divergence in Antirrhinum and the importance of geographic barriers.

Here I present genomic evidence showing the importance of factors like geographic barriers and environmental preferences in the divergence of the genus *Antirrhinum* at different scales. At the genus level (chapter 2), I found a strong geographic influence on the topology of the phylogenetic tree with populations from the Sierra Nevada and Morocco more closely related to each other than to other conspecific samples. This result was later supported by the analysis of genetic structure at the population level (chapter 3), where I found a strong relationship between latitude and longitude with the genetic clusters present in the PCA. Additionally, my results show the role of the Sierra Nevada as a barrier to gene flow between populations distributed in the north and south of this mountain range.

Although I found that geographic barriers (e.g. mountain ranges) are important for the genetic structure in *Antirrhinum*, the relationship between geography and the genetic variation of populations is also likely to be a consequence of a range of biotic and abiotic factors affecting migration. One of these factors is pollen dispersal. In *Antirrhinum*, most of the species have a predominant multiallelic self-incompatibility system which prevents self-fertilisation. This means that any pollen carrying a specific S-allele matching the one in the pistil will be rejected preventing fertilization (Vieira and Charlesworth, 2002). The combination of this mating system and the presence of an occluded corolla hiding the anthers from the exterior, makes pollinators essential for *Antirrhinum* reproduction and therefore for the maintenance of the genetic diversity of populations (Carrió and Güemes, 2014; Vargas *et al*., 2017). In a previous study, Vargas *et al*. (2017) characterised the pollinators of 18 different taxa of *Antirrhinum* in the Iberian Peninsula. They found that species distributed in the south of Spain were pollinated by a few species of bees with a common territorial behaviour. An example of how pollinator behaviour can affect the genetic structure of *Antirrhinum* was discussed by Torres *et al*. (2003) in the species *A. microphyllum*. This species distributed in the centre of Spain is reported as having a short distance genetic structuring which is likely related to the

behaviour of its main pollinator, the bee species *Rhodanthidium sticticum*. The territorial behaviour of this insect means males patrol small areas of a few square meters while the females collect pollen from patches within a 10m distance. Therefore is likely that some of the geographic variation in genetic structure in *Antirrhinum* will be explained by pollinator distribution and behaviour.

Another important factor in the genetic structure in plant populations is seed dispersal. In *Antirrhinum,* it is widely accepted that this occurs by a combination of gravity and wind, with seeds not falling far from the parental plants (Torres *et al*., 2003; Carrió *et al*., 2009). This method of seed dispersal is traditionally considered to promote differentiation between populations by increasing the likelihood of inbreeding within populations (Loveless and Hamrick, 1984). However, the strong self-incompatibility system in most *Antirrhinum* taxa will reduce the inbreeding in great extent in most species. Therefore I consider that this factor alone will not have as much influence as pollinators and pollen dispersal in the evolution of *Antirrhinum* populations.

Another important factor is the presence of suitable environmental conditions for the germination and establishment of *Antirrhinum* seeds. Wilson and Hudson (Unpublished) report a strong a relationship between the morphology of the main subsections and the environmental variables slope, substrate type and vegetation cover. Plants with *Kickxiella* morphologies are found on sites with rocky surfaces, high slopes and low vegetation cover while plants with *Antirrhinum* morphologies prefer level soils with higher vegetation cover. Torres *et al*. (2003), argues that the genetic structure in *A. microphyllum* (subsection *Kickxiella*) is related to the availability of cliffs and crevices necessary for the anchorage of roots. Therefore an uneven distribution of suitable environmental conditions across the landscape in combination with a limited dispersal of seeds could be the cause of the patterns of genetic structure found in this thesis.

Indirect evidence for the role of natural selection via habitat preferences in the divergence of *Antirrhinum* could be observed in the genomic structure of individuals

in hybrid populations from the south of the Sierra Nevada (chapter 4). I found a genome wide distribution of QTL involved in important adaptive traits. This suggests selection is acting over multiple traits that could be adaptive to a particular habitat (i.e. in dry environments several traits like leaf size and shape can subject to selection to reduce transpiration). However the role of natural selection in the morphology of *Antirrhinum* is discussed in more detail in the sections that follow.

Evidence of hybridisation in the evolution of Antirrhinum

My results supports the role of hybridisation as a key process in the evolution of *Antirrhinum* at different geographical and temporal scales. At the whole genus level (chapter 2), I found evidence of hybridisation between subsections *Kickxiella* and *Antirrhinum*. I also found signs of admixture in almost all *Kickxiella* species, with the only exceptions of *A. microphyllum*, *A. pertegasii*, and *A. valentinum*. These species have been reported as local endemics from east and central Spain (Wilson and Hudson, 2011), hence a limited distribution could explain a lack of admixture with members of subsection *Antirrhinum*. However, the fact that I could not detect admixture between subsections in these samples, does not rule out the possibility of hybridisation occurring between these taxa and other members of subsection *Kickxiella*.

Within subsection *Antirrhinum*, the species *A. latifolium* and *A. siculum* have equal percentages of ancestry from both subsections. These species are distributed in the north and south of Italy respectively and currently do not co-occur with any species from subsection *Kickxiella*. Therefore this could reflect old events of hybridisation with members of subsection *Kickxiella* at some point of the evolutionary history of the genus. On the other side, it has been reported that at least one of these species hybridizes with other members of subsection *Antirrhinum*. Güemes (2009) reports that *A. latifolium* distribution reaches the north of the Pyrenees where it hybridizes with the taxa *A. pseudomajus*. He even suggests that the taxa *A. striatum*, normally classified as a subspecies of the model system *A. majus*, is the result of this

hybridisation event. In this work I found that *A. striatum* is more related to *A. pseudomajus* (also considered a subspecies of *A. majus*) supporting its classification as a subspecies of *A. majus*. Nevertheless, the STRUCTURE analysis with values of K=3 shows an individual of *A. pseudomajus* having a small percentage of admixture from the *A. latifolium*/*A. siculum* genetic cluster supporting hybridisation between these taxa.

Evidence of hybridisation between subsections was also found in the population study between the taxa *A. barrelieri* and *A. rupestre* (chapter 3). By analysing the genetic structure of natural populations, I found a high proportion of advanced generation hybrids in the south of the Sierra Nevada. Similar results have been reported in another hybrid zone between two *Antirrhinum* taxa in the Pyrenees with divergent flower morphologies (Whibley *et al*., 2006; Bradley *et al*., 2017; Tavares *et al*., 2018). Although is very likely that hybridisation is a common process across the whole genus, as supported by my analysis with HyDe, hybridisation is a complex process that will depend on the interplay between intrinsic and extrinsic factors between species in contact. Therefore the outcome of these two independent hybridisation events in the Sierra Nevada and Pyrenees cannot be extrapolated to the rest of the genus. For example, no evidence of postzygotic barriers has been found between species in the hybrid populations at the Sierra Nevada and the Pyrenees (Andalo *et al*., 2010; Wilson *et al*., 2016). However, Carrió (2014) reports the existence of strong prezygotic barriers like divergent flowering times and genetic incompatibilities, between the species *A. barrelieri* and *A. valentinum* in the south of Spain.

My results also showed that most individuals in the south of the Sierra Nevada are advanced generation hybrids, providing evidence of constant gene flow between species. This result, in addition to the absence of intermediate morphologies and the rarity of pure individuals in these populations, raised questions about the maintenance of species boundaries and the genetic architecture of species differences. To understand more about these questions I compared my results from

the QTL analysis in an artificial hybrid population derived from species in the Sierra Nevada, with the results of a previous QTL study using the species *A. charidemi* and *A. majus* (Feng *et al*., 2009). *A. charidemi* and *A. majus* belong to subsection *Antirrhinum* and *Kickxiella* respectively and do not overlap in any of their natural ranges, hence I expected that the results from this comparison would show several important genomic features for the maintenance of morphological divergence in sympatric versus allopatric populations. My results showed 1) few QTL matching between studies. 2) A reduction in the number of QTL involved in each trait in the Sierra Nevada populations compared to the cross between *A. charidemi* and *A. majus*. 3) In the Sierra Nevada populations, an increase in the effect size of each QTL and 4) similar amounts of percentage of variance explained despite the great difference in number of QTL detected. Overall, the first two results correspond to the expected reduction in genetic diversity under the homogenising effects of gene flow in hybrid populations. However, the increase in individual effect sizes and the genome wide distribution of these loci, support the role of selection in the maintenance of morphological diversity in hybrid populations.

Morphological evolution

In this thesis I investigated the evolution of morphological traits in the genus *Antirrhinum*, a plant model system for the study of development and evolution. At the whole genus level, I found genetic evidence supporting two major genetic groups that broadly correspond to the way that the genus *Antirrhinum* is traditionally classified (*Antirrhinum* and *Kickxiella*). It is likely that these groups reflect two distinct evolutionary lineages that diverged early in the evolutionary history of the genus. However, my results also show that the genetic clusters do not correspond perfectly with the taxonomy. In particular, subsection *Kickxiella* appears in several places in the tree.

This parallel evolution could have happened in several ways. First, from standing genetic variation, i.e. by the fixation of polymorphic loci already present in the

ancestral population previous to the divergence of the genus. For example, the repeated selection of *Eda* alleles in the evolution of the armour plate in fresh water stickleback fish (Colosimo *et al*., 2005). Second, by the occurrence of independent mutations on different species or populations. For example, the independent evolution of the ability to feed from poisonous milkweeds in different insect species (Zhen *et al*., 2012). Third, by the introgression of adaptive alleles from one species to the other. For example, the evolution of colour patterns in different species of *Heliconius* butterflies (Dasmahapatra *et al*., 2012).

In *Antirrhinum*, results from the QTL analysis showed that adaptive phenotypic traits are controlled by QTL distributed throughout the genome (chapter 4) making it very unlikely that the origin of similar morphologies occurred via multiple independent mutations in all of these loci. In contrast, evidence supporting hybridisation between subsections suggests the occurrence of introgression of adaptive alleles from one subsection to the other. However, I did not find evidence of hybridisation or relatedness between species from the core *Kickxiella* clade and the species with *Kickxiella* like morphologies in the south of Spain (chapter 2). Although this does not exclude the possibility of previously undetected events of introgression a more reasonable explanation would be the selection of alleles already present in the ancestral population before the radiation of the genus. This scenario is supported by the results found in the hybrid populations and QTL analysis showing a role of natural selection via habitat preferences in the maintenance of morphological divergence despite hybridisation.

Another piece of evidence for a role of selection in the parallel evolution of the genus comes from the ancestral state character reconstruction (chapter 2). The results support a *Kickxiella* like morphology as the ancestral state for the genus. I found that this was followed by multiple shifts to an *Antirrhinum* like morphology, first within subsection *Streptosepalum* and later in subsection *Antirrhinum*, as suggested by the distribution of morphological traits on the phylogenetic tree. This transitions between different morphologies become more frequent within subsection

Antirrhinum. Additionally, my results show that the radiation of this subsection was likely followed by a shift in flower colour from white to magenta, supporting the bias towards gains in yellow and magenta colour reported for the tribe *Antirrhineae* (Ellis and Field, 2016). This shift in flower colour is likely of great importance for the evolution of the genus as several studies support the discrimination of pollinators between flower colours in *Antirrhinum* and the preference of magenta over white (Whibley *et al*., 2006; Whitney *et al*., 2013; Bradley *et al*., 2017; Tavares *et al*., 2018). Therefore is reasonable to argue that the direction in the evolution of these morphologies has a strong adaptive component, making more likely the possibility of parallel evolution by the selection of different adaptive alleles.

Future questions

Was parallel evolution involved in the evolution of adaptive morphology in Antirrhinum?

In the second chapter I found evidence suggesting parallel evolution in a clade of species with *Kickxiella* morphologies nested within subsection *Antirrhinum* and distributed in the Sierra Nevada. However, another possibility is ancestral introgression between subsections *Kickxiella* and *Antirrhinum*. A promising approach to test these hypotheses would be to compare the phylogenetic history of loci under divergence against neutral loci. A model of parallel evolution assumes that loci responsible for adaptive traits and neutral loci will show a geographic pattern in the tree. Alternatively, an adaptive radiation followed by hybridisation, assumes that adaptive alleles and neutral loci will show a discordant evolutionary history, with adaptive loci clustered by morphology and neutral by geography (Roda *et al*., 2013; Welch and Jiggins, 2014). This could be tested with a dense sampling of species and a whole genome resequence data. Additional information could also be obtained from the alignment of any chloroplast marker present among the RAD data with the *Antirrhinum* reference plastid genome. This could be used to build phylogenetic networks like those showed in Vargas, *et al.* (2009) and Hudson and Wilson (2011).

Is the change in flower colour related to an increased diversification rate in Antirrhinum radiation?

Subsection *Antirrhinum* has the highest number of species from all the three subsections and the highest number of transitions in morphology (chapter 2). At the base of this clade I observed a transition in flower colour from white to yellow, and from white to magenta. Considering the importance of variation in flower colour in natural populations in *Antirrhinum* (Whibley *et al*., 2006; Shaffer and Thomson, 2007; Tavares *et al*., 2018) I could argue that this shift was of central importance in the rapid radiation of the genus. In order to test this hypothesis future studies should seek to identify changes in diversification rates in this clade and a correlation with

shifts in flower colour. The main difficulty with this approach would be to obtain a time calibrated phylogeny that allow the estimation of such diversification rates, as there is a lack of fossil evidence in *Antirrhinum*.

Which genomic regions contribute to the maintenance of species boundaries in the hybrid populations?

In chapter 4 and 5 I aimed to understand more about the maintenance of species barriers in wild hybrid populations between two morphologically distinct species. Hence I identified genomic regions responsible for important adaptive traits and characterised their genomic architecture by performing a quantitative trait loci analysis (QTL). However, an issue with mapping methods is the possibility of not detecting loci with small effects or the possibility of not including an important trait underlying divergence in the analysis (e.g. flower colour). A complimentary approach to deal with this issues is to perform a genome scan for the detection of Fst outliers (Wright, 1931, 1946) as measure of relative divergence between populations. I addition, a link between Fst and the location of the QTL in this work would help to test the hypothesis of divergent selection driving morphological divergence in these species (Via, 2009). Preliminary analyses have been done as part of this study to answer this question, but the marker density used is too low to detect reliable regions of divergence.

Concluding remarks

In this work I investigate the genomics of divergence and hybridisation in the genus *Antirrhinum*. I show that geography has a strong influence in the phylogenetic relationships of the genus and found evidence of hybridisation between the main taxonomic subsections (chapter 2). However I also found difference between the genetic clustering and the traditional taxonomy of the genus, with subsection *Kickxiella* appearing in several places in the tree. I recovered a *Kickxiella* like morphology as the ancestral state of the genus, and found several transition events to an *Antirrhinum* like morphology throughout the genus.

In the third chapter I detect the presence of a high proportion of hybrids between the species *A. barrelieri* and *A. rupestre* in populations from the south of the Sierra Nevada, Spain. These individuals are advance generation hybrids and their morphology closely resembles the morphology of one the parental taxa. I also confirm the importance of geographic barriers in the evolution of the genus as I found that in *Antirrhinum* populations from the south of Spain, the Sierra Nevada represents an important barrier for gene flow.

Finally, in the fourth chapter I identify important genomic regions for the morphological divergence of species between subsection *Kickxiella* and *Antirrhinum*. I report that adaptive traits are controlled by several genome wide distributed QTL with different effect sizes and find regions with multiple overlapping QTL. After comparing my results with a previous QTL analysis in a pair of species without a hybridisation history I found a reduction in the number of QTL involved per trait in the populations from the Sierra Nevada likely caused by the homogenising effects of gene flow. The identification of matching regions bearing multiple QTL in both studies could suggest these sites are important for the divergence of the morphological subsections.

References

ABBOTT, R., D. ALBACH, S. ANSELL, J.W. ARNTZEN, S.J.E. BAIRD, N. BIERNE, J. BOUGHMAN, ET AL. 2013. Hybridization and speciation. *Journal of Evolutionary Biology*.

ABBOTT, R.J. 2017. Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution*.

ABBOTT, R.J., N.H. BARTON, and J.M. GOOD. 2016. Genomics of hybridization and its evolutionary consequences. *Molecular ecology*.

ADEYANJU, M.A. 2003. Phylogenetic analysis of the genus Antirrhinum based on the epidermal characters and the FIM-ERG intergenetic sequences. The University of Edinburgh.

ANASTASIOU, E., and M. LENHARD. 2007. Growing up to one's standard. *Current Opinion in Plant Biology*.

ANDALO, C., M.B. CRUZAN, C. CAZETTES, B. PUJOL, M. BURRUS, and C. THÉBAUD. 2010. Post-pollination barriers do not explain the persistence of two distinct Antirrhinum subspecies with parapatric distribution. *Plant Systematics and Evolution*.

BAACK, E.J., and L.H. RIESEBERG. 2007. A genomic view of introgression and hybrid speciation. *Current opinion in genetics & development* 17: 513–518.

BALDWIN, B.G., and M.J. SANDERSON. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proceedings of the National Academy of Sciences*.

BARTON, N. 1998. The geometry of adaptation. *Nature*.

BEAVIS, W.D., and BEAVIS. 1994. Beavis, W.D., 1994 The power and deceit of QTL experiments: Lessons from comparative QTL studies. *Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference*.

BLISCHAK, P.D., J. CHIFMAN, A.D. WOLFE, and L.S. KUBATKO. 2018. HyDe: a Python package for genome-scale hybridization detection. *bioRxiv*. Available at: http://biorxiv.org/content/early/2018/01/22/188037.abstract.

BOHONAK, A.J. 1999. Dispersal, gene flow, and population structure. *Quarterly review of biology* 74: 21–45.

BOLGER, A.M., M. LOHSE, and B. USADEL. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–20.

BOOKER, T., R.W. NESS, and D. CHARLESWORTH. 2015. Molecular Evolution: Breakthroughs and Mysteries in Batesian Mimicry. *Current Biology*.

BRADBURY, P.J., Z. ZHANG, D.E. KROON, and T.M. CASSTEVENS. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 3: 2633–2635.

BRADLEY, D., P. XU, I.-I. MOHORIANU, A. WHIBLEY, D. FIELD, H. TAVARES, M. COUCHMAN, ET AL. 2017. Evolution of flower color pattern through selection on regulatory small RNAs. *Science* 358: 925 LP – 928. Available at: http://science.sciencemag.org/content/358/6365/925.abstract.

BRAWAND, D., C.E. WAGNER, Y.I. LI, M. MALINSKY, I. KELLER, S. FAN, O. SIMAKOV, ET AL. 2015. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*.

BRENNAN, A.C., S.J. HISCOCK, and R.J. ABBOTT. 2016. Genomic architecture of phenotypic divergence between two hybridizing plant species along an elevational gradient. *AoB PLANTS*.

BROAD INSTITUTE. 2018. Picard tools. Available at: http://broadinstitute.github.io/picard/.

BROMAN, K.W., and S. SEN. 2009. A Guide to QTL Mapping with R/qtl. Springer, New York.

BROMAN, K.W., H. WU, Ś. SEN, and G.A. CHURCHILL. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*19:889-890.

CAMPBELL, C.R., J.W. POELSTRA, and A.D. YODER. 2018. What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*.

CARRIÓ, E., and J. GÜEMES. 2014. The effectiveness of pre- and post-zygotic barriers in avoiding hybridization between two snapdragons (Antirrhinum L.: Plantaginaceae). *Botanical Journal of the Linnean Society* 176: 159–172. Available at: http://dx.doi.org/10.1111/boj.12201.

CARRIÓ, E., J.F. JIMÉNEZ, P. SÁNCHEZ-GÓMEZ, and J. GÜEMES. 2009. Reproductive biology and conservation implications of three endangered snapdragon species (Antirrhinum, Plantaginaceae). *Biological Conservation*.

CATCHEN, J., P.A. HOHENLOHE, S. BASSHAM, A. AMORES, and W.A. CRESKO. 2013. Stacks: an analysis tool set for population genomics. *Molecular ecology* 22: 3124–3140.

CAVENDER-BARES, J., A. GONZÁLEZ-RODRÍGUEZ, D.A.R. EATON, A.A.L. HIPP, A. BEULKE, and P.S. MANOS. 2015. Phylogeny and biogeography of the american live oaks (Quercus subsection Virentes): A genomic and population genetics approach. *Molecular Ecology*.

CHAFFE, C.J. 2003. Molecular and genetic characterisation of the European snapdragons (Scrophulariaceae: Antirrhinum spp.). Univeristy of East Anglia.

CHAN, Y.F., M.E. MARKS, F.C. JONES, G. VILLARREAL, M.D. SHAPIRO, S.D. BRADY, A.M. SOUTHWICK, ET AL. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a pitxl enhancer. *Science*.

CHIFMAN, J., and L. KUBATKO. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of theoretical biology* 374: 35–47.

CHIFMAN, J., and L. KUBATKO. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics*.

COEN, E.S., and E.M. MEYEROWITZ. 1991. The war of the whorls: Genetic interactions controlling flower development. *Nature*.

COLLARD, B.C.Y., M.Z.Z. JAHUFER, J.B. BROUWER, and E.C.K. PANG. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*.

COLOSIMO, P.F., K.E. HOSEMANN, S. BALABHADRA, G. VILLARREAL, H. DICKSON, J. GRIMWOOD, J. SCHMUTZ, ET AL. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*.

COMES, H.P., and J.W. KADEREIT. 1998. The effect of quaternary climatic changes on

plant distribution and evolution. *Trends in Plant Science*.

COYNE, J.A., and H.A. ORR. 2004. Speciation. Sinauer Associates Inc., U.S.A.

CRAWFORD, B.C.W. 2004. CINCINNATA Controls Both Cell Differentiation and Growth in Petal Lobes and Leaves of Antirrhinum. *PLANT PHYSIOLOGY*.

CRUICKSHANK, T.E., and M.W. HAHN. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*.

CRUZAN, M.B., and M.L. ARNOLD. 1993. Ecological and Genetic Associations in an Iris Hybrid Zone. *Evolution*.

DANECEK, P., A. AUTON, G. ABECASIS, C.A. ALBERS, E. BANKS, M.A. DEPRISTO, R.E. HANDSAKER, ET AL. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158. Available at: http://dx.doi.org/10.1093/bioinformatics/btr330.

DASMAHAPATRA, K.K., J.R. WALTERS, A.D. BRISCOE, J.W. DAVEY, A. WHIBLEY, N.J. NADEAU, A. V. ZIMIN, ET AL. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94–98.

DRUMMOND, C.S., R.J. EASTWOOD, S.T.S. MIOTTO, and C.E. HUGHES. 2012. Multiple continental radiations and correlates of diversification in lupinus (leguminosae): Testing for key innovation with incomplete taxon sampling. *Systematic Biology*.

ELLIS, T.J., and D.L. FIELD. 2016. Repeated gains in yellow and anthocyanin pigmentation in flower colour transitions in the Antirrhineae. *Annals of Botany*.

EMERSON, K.J., C.R. MERZ, J.M. CATCHEN, P.A. HOHENLOHE, W.A. CRESKO, W.E. BRADSHAW, and C.M. HOLZAPFEL. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the national academy of sciences* 107: 16196–16200.

EVANNO, G., S. REGNAUT, and J. GOUDET. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*.

FEDER, J.L., S.P. EGAN, and P. NOSIL. 2012. The genomics of speciation-with-gene-flow.

Trends in Genetics 28: 342–350.

FENG, X., Y. WILSON, J. BOWERS, R. KENNAWAY, A. BANGHAM, A. HANNAH, E. COEN, and A. HUDSON. 2009. Evolution of allometry in Antirrhinum. *The Plant Cell* 21: 2999–3007.

FERRIS, K.G., L.L. BARNETT, B.K. BLACKMAN, and J.H. WILLIS. 2017. The genetic architecture of local adaptation and reproductive isolation in sympatry within the Mimulus guttatus species complex. *Molecular Ecology*.

FITZJOHN, R.G. 2012. Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*.

GARCÍA-BARRIUSO, M., A.L. CRESPÍ, C. NABAIS, S. BERNARDOS, and F. AMICH. 2012. Mineral element composition in Antirrhinum subsection Streptosepalum (Plantaginaceae) in Western Europe (Iberian Peninsula). *Lazaroa*.

GARRISON, E. 2018. vcflib: A C++ library for parsing and manipulating VCF files. Available at: https://github.com/vcflib/vcflib [Accessed June 16, 2018].

GIVNISH, T.J. 2015. Adaptive radiation versus "radiation" and "explosive diversification": Why conceptual distinctions are fundamental to understanding evolution. *New Phytologist*.

GLOR, R.E. 2010. Phylogenetic Insights on Adaptive Radiation. *Annual Review of Ecology, Evolution, and Systematics*.

GLOVER, B. 2014. Understanding Flowers and Flowering Second Edition. 2nd editio. Oxford University Press.

GRISWOLD, C.K. 2006. Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity*.

GUBITZ, T., A. CALDWELL, and A. HUDSON. 2003. Rapid molecular evolution of CYCLOIDEA-like genes in Antirrhinum and its relatives. *Molecular biology and evolution* 20: 1537–1544.

GÜEMES, J. 2009. Antirrhinum L. *In* S. Castroviejo, A. Herrero, C. Benedí, E. Rico, and J. Güemes [eds.], Flora Iberica, 134–166. CSIC, Madrid.

HANNA, A.I. 2007. AAMToolbox user guide.

HARRISON, R.G., and E.L. LARSON. 2014. Hybridization, introgression, and the nature of species boundaries. *In* Journal of Heredity,

HASEGAWA, M., H. KISHINO, and T. AKI YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*.

HAWTHORNE, D.J., and S. VIA. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412: 904–907.

HEGARTY, M.J., and S.J. HISCOCK. 2005. Hybrid speciation in plants: New insights from molecular studies. *New Phytologist*.

VAN HENGSTUM, T., S. LACHMUTH, J.G.B. OOSTERMEIJER, J.C.M. DEN NIJS, P.G. MEIRMANS, and P.H. VAN TIENDEREN. 2012. Human-induced hybridization among congeneric endemic plants on Tenerife, Canary Islands. *Plant Systematics and Evolution*.

HERSCH-GREEN, E.I., and R. CRONN. 2009. Tangled trios?: Characterizing a hybrid zone in Castilleja (Orobanchaceae). *American Journal of Botany*.

HEWITT, G.M. 1988. Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology and Evolution*.

HEWITT, G.M. 1999. Post-glacial re-colonization of European biota. *In* Biological Journal of the Linnean Society,

HEYWOOD, J. 1986. Clinal variation associated with edaphic ecotones in hybrid populations of Gaillardia pulchella. *Evolution*.

HOBAN, S.M., T.S. MCCLEARY, S.E. SCHLARBAUM, S.L. ANAGNOSTAKIS, and J. ROMERO-SEVERSON. 2012. Human-impacted landscapes facilitate hybridization between a native and an introduced tree. *Evolutionary Applications*.

HODGES, S.A., and N.J. DERIEG. 2009. Adaptive radiations: From field to genomic studies. *Proceedings of the National Academy of Sciences*.

HUBER, B., A. WHIBLEY, Y.L. POUL, N. NAVARRO, A. MARTIN, S. BAXTER, A. SHAH, ET AL. 2015. Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies. *Heredity*.

HUDSON, A., J. CRITCHLEY, and Y. ERASMUS. 2008a. Propagating Antirrhinum. *Cold Spring Harbor Protocols* 2008: .

HUDSON, A., J. CRITCHLEY, and Y. ERASMUS. 2007. The genus antirrhinum (snapdragon): a flowering plant model for evolution and development. *The genus antirrhinum (snapdragon): a flowering plant model for evolution and development.* 3: .

HUDSON, A., J. CRITCHLEY, and Y. ERASMUS. 2008b. The genus Antirrhinum (snapdragon): A flowering plant model for evolution and development. *Cold Spring Harbor Protocols*.

HUGHES, C.E., R. NYFFELER, and H.P. LINDER. 2015. Evolutionary plant radiations: Where, when, why and how? *New Phytologist*.

HUSON, D.H. 1998. SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics*.

INGRAM, G.C., and R. WAITES. 2006. Keeping it together: Co-ordinating plant growth. *Current Opinion in Plant Biology*.

JIMÉNEZ, J.F., P. SÁNCHEZ-GÓMEZ, and J. GÜEMES. 2005. Phylogeny of snapdragon species (Antirrhinum; Scrophulariaceae) using non-coding cpDNA sequences. *Phylogeny of snapdragon species (Antirrhinum; Scrophulariaceae) using non-coding cpDNA sequences*.

JONES, J.C., S. FAN, P. FRANCHINI, M. SCHARTL, and A. MEYER. 2013. The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology* 22: 2986–3001.

KAHLE, D., and H. WICKHAM. 2013. ggmap: Spatial Visualization with ggplot2. *R Journal* 5: .

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press.

DE LA HARPE, M., M. PARIS, D.N. KARGER, J. ROLLAND, M. KESSLER, N. SALAMIN, and C. LEXER. 2017. Molecular ecology studies of species radiations: Current research gaps, opportunities and challenges. *Molecular Ecology*.

LANDER, E.S., and S. BOTSTEIN. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*.

LANGLADE, N.B., X. FENG, T. DRANSFIELD, L. COPSEY, A.I. HANNA, C. THÉBAUD, A. BANGHAM, ET AL. 2005. Evolution through genetically controlled allometry space. *Proceedings of the National Academy of Sciences of the united states of america* 102: 10221–10226.

LANGMEAD, B., and S.L. SALZBERG. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357. Available at: http://dx.doi.org/10.1038/nmeth.1923.

LARTER, M., A. DUNBAR-WALLIS, A.E. BERARDI, and S.D. SMITH. 2018. Convergent evolution at the pathway level: Predictable regulatory changes during flower color transitions. *Molecular Biology and Evolution*.

LENORMAND, T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* 17: .

LI, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)* 27: 2987–2993.

LI, M., D. ZHANG, Q. GAO, Y. LUO, H. ZHANG, B. MA, C. CHEN, ET AL. 2019. Genome structure and evolution of Antirrhinum majus L. *Nature Plants*.

LI, M., D. ZHANG, H. ZHANG, Q. GAO, B. MA, C. CHEN, Y. LUO, ET AL. 2018. Genome structure and evolution of Antirrhnum majus L. *bioRxiv*.

LIBERAL, I.M., M. BURRUS, C. SUCHET, C. THÉBAUD, and P. VARGAS. 2014. The evolutionary history of Antirrhinum in the Pyrenees inferred from phylogeographic analyses. *BMC Evolutionary Biology* 14: 146.

LINDTKE, D., S.C. GONZÁLEZ-MARTÍNEZ, D. MACAYA-SANZ, and C. LEXER. 2013. Admixture mapping of quantitative traits in Populus hybrid zones: Power and limitations. *Heredity*.

LIU, C.-W., and J. MURRAY. 2016. The Role of Flavonoids in Nodulation Host-Range Specificity: An Update. *Plants*.

LOUTHAN, A.M., and K.M. KAY. 2011. Comparing the adaptive landscape across trait types: Larger QTL effect size in traits under biotic selection. *BMC Evolutionary Biology*.

LOVELESS, M.D., and J.L. HAMRICK. 1984. Ecological Determinants of Genetic Structure in Plant Populations. *Annual Review of Ecology and Systematics*.

LUO, D., R. CARPENTER, C. VINCENT, L. COPSEY, and E. COEN. 1996. Origin of floral asymmetry in Antirrhinum. *Nature*.

MALEK, T.B., J.W. BOUGHMAN, I. DWORKIN, and C.L. PEICHEL. 2012. Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Molecular Ecology*.

MALLET, J. 2007. Hybrid speciation. *Nature* 446: 279–283.

MARQUES, D.A., K. LUCEK, J.I. MEIER, S. MWAIKO, C.E. WAGNER, L. EXCOFFIER, and O. SEEHAUSEN. 2016. Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLoS Genetics*.

MARTIN, S.H., K.K. DASMAHAPATRA, N.J. NADEAU, C. SALAZAR, J.R. WALTERS, F. SIMPSON, M. BLAXTER, ET AL. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research*.

MINDER, A.M., C. ROTHENBUEHLER, and A. WIDMER. 2007. Genetic structure of hybrid zones between Silene latifolia and Silene dioica (Caryophyllaceae): Evidence for introgressive hybridization. *Molecular Ecology*.

MIZUKAMI, Y. 2001. A matter of size: Developmental control of organ size in plants. *Current Opinion in Plant Biology*.

NADEAU, N.J., S.H. MARTIN, K.M. KOZAK, C. SALAZAR, K.K. DASMAHAPATRA, J.W. DAVEY, S.W. BAXTER, ET AL. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology* 22: 814–826.

NEIGE, P. 2015. 1 – A Singular Work of Theater. *In* F. Gaill [ed.], Events of Increased Biodiversity, 1–12. Elsevier, Oxford.

NG, J., L.B. FREITAS, and S.D. SMITH. 2018. Stepwise evolution of floral pigmentation predicted by biochemical pathway structure. *Evolution*.

NOOR, M.A.F., and J.L. FEDER. 2006. Speciation genetics: Evolving approaches. *Nature Reviews Genetics*.

NOSIL, P., and A. BUERKLE. 2010. Population Genomics. *Nature Education Knowledge* 3:

8.

NOSIL, P., D.J. FUNK, and D. ORTIZ-BARRIENTOS. 2009. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*.

NOSIL, P., T.L. PARCHMAN, J.L. FEDER, and Z. GOMPERT. 2012. Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in Timema stick insects. *BMC Evolutionary Biology*.

OMLAND, K.E. 1999. The assumptions and challenges of ancestral state reconstructions. *Systematic Biology*.

ORR, H.A. 2005. The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*.

OUELLETTE, L.A., R.W. REID, S.G. BLANCHARD, and C.R. BROUWER. 2018. LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics*.

PETREN, K., P.R. GRANT, B.R. GRANT, and L.F. KELLER. 2005. Comparative landscape genetics and the adaptive radiation of Darwin's finches: The role of peripheral isolation. *Molecular Ecology*.

PRICE, N., B.T. MOYERS, L. LOPEZ, J.R. LASKY, J.G. MONROE, J.L. MULLEN, C.G. OAKLEY, ET AL. 2018. Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*.

PRITCHARD, J.K., M. STEPHENS, and P. DONNELLY. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945–959.

PRUNIER, R., and K.E. HOLSINGER. 2010. Was it an explosion? Using population genetics to explore the dynamics of a recent radiation within Protea (Proteaceae L.). *Molecular Ecology*.

RAGLAND, G.J., M.M. DOELLMAN, P.J. MEYERS, G.R. HOOD, S.P. EGAN, T.H.Q. POWELL, D.A. HAHN, ET AL. 2017. A test of genomic modularity among life-history adaptations promoting speciation with gene flow. *Molecular Ecology*.

RAJ, A., M. STEPHENS, and J.K. PRITCHARD. 2014. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*.

RAMBAUT, A. 2016. FigTree v1.4.3. *Molecular evolution, phylogenet ics and epidemiology*.

RAUSHER, M.D. 2008. Evolutionary Transitions in Floral Color. *International Journal of Plant Sciences*.

RAVINET, M., R. FARIA, R.K. BUTLIN, J. GALINDO, N. BIERNE, M. RAFAJLOVIĆ, M.A.F. NOOR, ET AL. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*.

RENAUT, S., G.L. OWENS, and L.H. RIESEBERG. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology*.

REVELL, L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*.

RICHARDSON, J.E., R.T. PENNINGTON, T.D. PENNINGTON, and P.M. HOLLINGSWORTH. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science*.

RICHARDSON, J.E., F.M. WEITZ, M.F. FAY, Q.C.B. CRONK, H.P. LINDER, G. REEVES, and M.W. CHASE. 2001. Rapid and recent origin of species richness in the Cape flora of south Africa. *Nature*.

RIESEBERG, L., and C.A. BUERKLE. 2002. Genetic mapping in hybrid zones. *The American naturalist*.

RIESEBERG, L.H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics*.

RIESEBERG, L.H., S.J.E. BAIRD, and K.A. GARDNER. 2000. Hybridization, introgression, and linkage evolution. *Plant Molecular Biology*.

RIESEBERG, L.H., and B.K. BLACKMAN. 2010. Speciation genes in plants. *Annals of Botany*.

RIESEBERG, L.H., C. FOSSEN, and A.M. DESROCHERS. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375: 313–316.

RIESEBERG, L.H., and J.H. WILLIS. 2007. Plant speciation. *Science*.

RODA, F., L. AMBROSE, G.M. WALTER, H.L. LIU, A. SCHAUL, A. LOWE, P.B. PELSER, ET AL. 2013.

Genomic evidence for the parallel evolution of coastal forms in the Senecio lautus complex. *Molecular Ecology*.

ROTHMALER, W. 1956. Taxonomische monographie Antirrhinum. Akademie Verlag, Berlin.

RUBIN, B.E.R., R.H. REE, and C.S. MOREAU. 2012. Inferring Phylogenies from RAD Sequence Data. *PLoS ONE* 7: 1–12.

SCHLUTER, D. 2009. Evidence for Ecological Speciation and Its Alternative. *Science* 323: 737–741. Available at: http://www.sciencemag.org/content/323/5915/737.abstract.

SCHWARZ-SOMMER, Z., B. DAVIES, and A. HUDSON. 2003. An everlasting pioneer: the story of Antirrhinum research. *Nature reviews. Genetics* 4: 657–666.

SCHWARZ-SOMMER, Z., E.D.A. SILVA, R. BERNDTGEN, W. LO, A. MU, I. NINDL, K. STU, ET AL. 2003. A Linkage Map of an F 2 Hybrid Population of Antirrhinum majus and A . molle. *Genetics* 710: 699–710.

SHAFFER, H.B., and R.C. THOMSON. 2007. Delimiting species in recent radiations. *Systematic Biology*.

SHANG, Y., J. VENAIL, S. MACKAY, P.C. BAILEY, K.E. SCHWINN, P.E. JAMESON, C.R. MARTIN, and K.M. DAVIES. 2011. The molecular basis for venation patterning of pigmentation and its effect on pollinator attraction in flowers of Antirrhinum. *New Phytologist*.

SLATE, J. 2013. From beavis to beak color: A simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. *Evolution*.

SLATKIN, M. 1987. Gene flow and the geographic structure of natural populations. *Science*.

SMITH, S.D., and E.E. GOLDBERG. 2015. Tempo and mode of flower color evolution. *American Journal of Botany*.

SMITH, S.L., R.F. CARDEN, B. COAD, T. BIRKITT, and J.M. PEMBERTON. 2014. A survey of the hybridisation status of Cervus deer species on the island of Ireland. *Conservation Genetics*.

SOLTIS, P.S., and D.E. SOLTIS. 2009. The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology*.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. Available at: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu033 [Accessed April 24, 2018].

STEBBINS, G.L. 1950. Variation and evolution in plants. Columbia University Press, New York.

STRAUSS, S.Y., and J.B. WHITALL. 2006. Non-pollinator agents of selection on floral traits. *In* Ecology and evolution of flowers, 120–138. Oxford University Press, Oxford, UK.

SUTTON, D.A. 1988. A revision of the tribe Antirrhineae. Oxford University Press, New York.

TAVARES, H., A. WHIBLEY, D.L. FIELD, D. BRADLEY, M. COUCHMAN, L. COPSEY, J. ELLEOUET, ET AL. 2018. Selection and gene flow shape genomic islands that control floral guides. *Proceedings of the National Academy of Sciences*.

TORRES, E., J.M. IRIONDO, A. ESCUDERO, and C. PÉREZ. 2003. Analysis of within-population spatial genetic structure in Antirrhinum microphyllum (Scrophulariaceae). *American Journal of Botany*.

TURNER, T.L., M.W. HAHN, and S. V. NUZHDIN. 2005. Genomic islands of speciation in Anopheles gambiae. *PLoS Biology*.

TUTIN, T.G., V.H. HEYWOOD, N.A. BURGES, D.M. MOORE, D.H. VALENTINE, S.M. WALTERS, and D.A. WEBB. 1972. Flora Europaea Volume 3: Diapensiaceae to Myoporaceae. Cambridge University Press, Cambridge.

VARGAS, P., E. CARRIÓ, B. GUZMÁN, E. AMAT, and J. GÜEMES. 2009. A geographical pattern of Antirrhinum (Scrophulariaceae) speciation since the Pliocene based on plastid and nuclear DNA polymorphisms. *Journal of Biogeography* 36: 1297–1312.

VARGAS, P., I. LIBERAL, C. ORNOSA, and J.M. GÓMEZ. 2017. Flower specialisation: the

occluded corolla of snapdragons (Antirrhinum) exhibits two pollinator niches of large long-tongued bees. *Plant Biology*.

VARGAS, P., J.A. ROSSELLÓ, R. OYAMA, and J. GÜEMES. 2004. Molecular evidence for naturalness of genera in the tribe Antirrhineae (Scrophulariaceae) and three independent evolutionary lineages from the New World and the Old. *Plant Systematics and Evolution*151–172.

VARGAS, P., L.M. VALENTE, J.L. BLANCO-PASTOR, I. LIBERAL, B. GUZMÁN, E. CANO, A. FORREST, and M. FERNÁNDEZ-MAZUECOS. 2013. Testing the biogeographical congruence of palaeofloras using molecular phylogenetics: Snapdragons and the Madrean-Tethyan flora. *Journal of Biogeography*.

VIA, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

VIA, S. 2009. Natural selection in action during speciation. *Proceedings of the National Academy of Sciences* 106: 9939–9946.

VIEIRA, C.P., and D. CHARLESWORTH. 2002. Molecular variation at the self-incompatibility locus in natural populations of the genera Antirrhinum and Misopates. *Heredity*.

WAGNER, C.E., I. KELLER, S. WITTWER, O.M. SELZ, S. MWAIKO, L. GREUTER, A. SIVASUNDAR, and O. SEEHAUSEN. 2013a. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *In* Molecular Ecology,

WAGNER, C.E., I. KELLER, S. WITTWER, O.M. SELZ, S. MWAIKO, L. GREUTER, A. SIVASUNDAR, and O. SEEHAUSEN. 2013b. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular ecology* 22: 787–798.

WEBB, D.A. 1971. Taxonomic notes in Antirrhinum. *Botanical Journal of the Linnean Society* 10: 271:274.

WELCH, J.J., and C.D. JIGGINS. 2014. Standing and flowing: The complex origins of

adaptive variation. *Molecular Ecology*.

WHIBLEY, A.C., N.B. LANGLADE, C. ANDALO, A.I. HANNA, A. BANGHAM, C. THÉBAUD, and E. COEN. 2006. Evolutionary paths underlying flower color variation in Antirrhinum. *Science*.

WHITNEY, H.M., G. MILNE, S.A. RANDS, S. VIGNOLINI, C. MARTIN, and B.J. GLOVER. 2013. The influence of pigmentation patterning on bumblebee foraging from flowers of Antirrhinum majus. *Naturwissenschaften*.

WHITNEY, K.D., J.R. AHERN, L.G. CAMPBELL, L.P. ALBERT, and M.S. KING. 2010. Patterns of hybridization in plants. *Perspectives in Plant Ecology, Evolution and Systematics*.

WILKINSON, L. 2011. ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics*.

WILSON, Y., J. CRITCHLEY, C. BOUVART, C. BROMFIELD, and A. HUDSON. 2016. Genetic evidence for ecological speciation in the genus Antirrhinum. Edinburgh.

WILSON, Y., and A. HUDSON. 2011. The evolutionary history of Antirrhinum suggests that ancestral phenotype combinations survived repeated hybridizations. *The Plant Journal* 66: 1032–1043.

WRIGHT, S. 1931. Evolution in Mendelian Populations. *Genetics* 16: 97–159.

WRIGHT, S. 1946. Isolation by distance under diverse systems of mating. *Genetics*.

WU, T.D., and S. NACU. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*.

YEAMAN, S., and S.P. OTTO. 2011. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*.

YEAMAN, S., and M.C. WHITLOCK. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution*.

ZHANG, D., Q. YANG, W. BAO, Y. ZHANG, B. HAN, Y. XUE, and Z. CHENG. 2005. Molecular cytogenetic characterization of the Antirrhinum majus genome. *Genetics* 169: 325–335.

ZHEN, Y., M.L. AARDEMA, E.M. MEDINA, M. SCHUMER, and P. ANDOLFATTO. 2012. Parallel molecular evolution in an herbivore community. *Science*.

Supplementary material

Table S.1. Geographic coordinates of the samples used for phylogenetic and STRUCTURE analyses.

Taxa	Population	Longitude	Latitude
A. australe	L84	-5.38207	36.76648
A. australe	L91	-5.27826	36.61832
A. australe	L91	-5.27826	36.61832
A. boissieri	L18	-3.40861	37.59884
A. boissieri	L18	-3.40861	37.59884
A. boissieri	L104	-4.42587	37.47293
A. cirrigherum	L114	-8.78155	40.42464
A. cirrigherum	L114	-8.78155	40.42464
A. graniticum	L67	-2.61528	41.01546
A. graniticum	L69	-2.84585	40.86696
A. graniticum	L69	-2.84585	40.86696
A. graniticum	L115	-7.79998	41.07849
A. graniticum	L119	-5.69698	40.25302
A. graniticum	L120	-5.29789	40.17781
A. grosii	L175	-5.27365	40.25633
A. graniticum	L115	-7.79998	41.07849
A. linkianum	L108	-9.16709	38.667
A. linkianum	L110	-8.68523	39.39877
A. linkianum	L111	-8.81163	39.59898
A. linkianum	L112	-8.47849	40.13557
A. litigiosum	L61	1.363061	41.36511
A. litigiosum	L64	-1.33322	41.83451
A. litigiosum	L66	-1.33	41.84
A. litigiosum	L80	-1.46667	41.05813
A. lopesianum	L74	-6.52	41.58
A. barrelieri_Morocco	L161	-5.44008	35.51007
A. barrelieri_Morocco	L161	-5.44008	35.51007
A. barrelieri_Morocco	L167	-4.19137	34.63283
A. barrelieri	L148	-3.314	36.94
A. barrelieri	L150	-3.183	36.9363
A. barrelieri	L205	-3.2373	36.7529
A. meonanthum	L118	-7.51581	40.37837
A. hispanicum_Morocco	L172	-5.51405	32.74138
A. hispanicum_Morocco	L172	-5.51405	32.74138
A. microphyllum	L72	-2.74929	40.49214
A. microphyllum	L73	-2.78444	40.39892

A. microphyllum	L73	-2.78444	40.39892
A. mollisimum	L19	-2.80437	37.1326
A. mollisimum	L20	-2.65733	36.86748

Table S.2. Geographic coordinates of the populations used in the population and hybrid analyses (chapter 2).

Taxa	Forward/Reverse	Longitude	Latitude
A.mollissimum	20	-2.657331	36.86748
A.barrelieri	96	-4.701603	36.799
A.barrelieri	131	-3.4923	36.8646
A.hispanicum	132	-3.471016667	36.92142
A.barrelieri	135	-3.28579	36.8983
A.rupestre	136	-3.262716667	36.99953
A.barrelieri	139	-3.348429377	36.95197
A.hispanicum	141	-3.43	36.914
A.barrelieri	142	-3.378	36.926
A.barrelieri	146	-3.2815	36.95433
A.barrelieri	148	-3.314	36.94
A.barrelieri	149	-3.216333333	36.95742
A.barrelieri	150	-3.183	36.9363
A.rupestre	152	-3.159	36.9765
A.barrelieri	154	-3.12	36.98
A.barrelieri	156	-3.009	37.0094
A.tortuosum	159	-4.1257	37.1644
A.hispanicum	201	-3.40094	36.8779
A.barrelieri	205	-3.2373	36.7529
A.hispanicum	206	-3.515567069	36.83344
A.hispanicum	207	-3.47828	37.1661
A.hispanicum	208	-3.21318	37.2706
A.hispanicum	209	-3.4213	37.3071

Table S.3. Number of paired reads per sample obtained from sequencing in chapter 1.

Sample	Forward/Reverse	P1 adaptor	P2 adaptor	Number of reads
A.australe_L84	Forward	TAATG	ATATC	176800
A.australe_L84	Reverse	TAATG	ATATC	176800
A.australe_L84	Forward	TAATG	CAGTC	93351
A.australe_L84	Reverse	TAATG	CAGTC	93351
A.australe_L91_1	Forward	GGTTC	ATATC	150746
A.australe_L91_1	Reverse	GGTTC	ATATC	150746
A.australe_L91_1	Forward	GGTTC	CAGTC	79009
A.australe_L91_1	Reverse	GGTTC	CAGTC	79009
A.australe_L91_2	Forward	ACCAT	ATATC	213142
A.australe_L91_2	Reverse	ACCAT	ATATC	213142
A.australe_L91_2	Forward	ACCAT	CAGTC	117413
A.australe_L91_2	Reverse	ACCAT	CAGTC	117413
A.barrelieri_L148	Forward	TCGAG	ATCGA	59976
A.barrelieri_L148	Reverse	TCGAG	ATCGA	59976
A.barrelieri_L148	Forward	TCGAG	CGTAT	363703
A.barrelieri_L148	Reverse	TCGAG	CGTAT	363703
A.barrelieri_L150_1	Forward	TGTGG	ATCGA	90618
A.barrelieri_L150_1	Reverse	TGTGG	ATCGA	90618
A.barrelieri_L150_1	Forward	TGTGG	CGTAT	564486
A.barrelieri_L150_1	Reverse	TGTGG	CGTAT	564486
A.barrelieri_L150_2	Forward	TTTTA	ATCGA	109695
A.barrelieri_L150_2	Reverse	TTTTA	ATCGA	109695
A.barrelieri_L150_2	Forward	TTTTA	CGTAT	718788
A.barrelieri_L150_2	Reverse	TTTTA	CGTAT	718788
A.barrelieri_L205_1	Forward	CAGTC	ATCGA	210410
A.barrelieri_L205_1	Reverse	CAGTC	ATCGA	210410
A.barrelieri_L205_1	Forward	CAGTC	CGTAT	1273368
A.barrelieri_L205_1	Reverse	CAGTC	CGTAT	1273368
A.barrelieri_L205_2	Forward	CCGGT	ATCGA	614752
A.barrelieri_L205_2	Reverse	CCGGT	ATCGA	614752
A.barrelieri_L205_2	Forward	CCGGT	CGTAT	3749423
A.barrelieri_L205_2	Reverse	CCGGT	CGTAT	3749423
A.barrelieri_Morocco_L161_1	Forward	AAAAA	GATCG	1899915
A.barrelieri_Morocco_L161_1	Reverse	AAAAA	GATCG	1899915
A.barrelieri_Morocco_L161_1	Forward	AAAAA	TCGAG	393324
A.barrelieri_Morocco_L161_1	Reverse	AAAAA	TCGAG	393324
A.barrelieri_Morocco_L161_2	Forward	GCGCC	GATCG	684062

A.barrelieri_Morocco_L161_2	Reverse	GCGCC	GATCG	684062
A.barrelieri_Morocco_L161_2	Forward	GCGCC	TCGAG	141216
A.barrelieri_Morocco_L161_2	Reverse	GCGCC	TCGAG	141216
A.barrelieri_Morocco_L167	Forward	CTAGG	GATCG	853136
A.barrelieri_Morocco_L167	Reverse	CTAGG	GATCG	853136
A.barrelieri_Morocco_L167	Forward	CTAGG	TCGAG	175113
A.barrelieri_Morocco_L167	Reverse	CTAGG	TCGAG	175113
A.boissieri_L104	Forward	GTCAC	ATATC	258767
A.boissieri_L104	Reverse	GTCAC	ATATC	258767
A.boissieri_L104	Forward	GTCAC	CAGTC	134164
A.boissieri_L104	Reverse	GTCAC	CAGTC	134164
A.boissieri_L18_1	Forward	ATCGA	ATATC	337687
A.boissieri_L18_1	Reverse	ATCGA	ATATC	337687
A.boissieri_L18_1	Forward	ATCGA	CAGTC	174022
A.boissieri_L18_1	Reverse	ATCGA	CAGTC	174022
A.boissieri_L18_2	Forward	AGCTG	ATATC	161244
A.boissieri_L18_2	Reverse	AGCTG	ATATC	161244
A.boissieri_L18_2	Forward	AGCTG	CAGTC	82732
A.boissieri_L18_2	Reverse	AGCTG	CAGTC	82732
A.braun-blanquetii_E20	Forward	AAAAA	ATATC	207214
A.braun-blanquetii_E20	Reverse	AAAAA	ATATC	207214
A.braun-blanquetii_E20	Forward	AAAAA	CAGTC	107293
A.braun-blanquetii_E20	Reverse	AAAAA	CAGTC	107293
A.charidemi_E23_1	Forward	GTTGT	ATATC	434258
A.charidemi_E23_1	Reverse	GTTGT	ATATC	434258
A.charidemi_E23_1	Forward	GTTGT	CAGTC	222192
A.charidemi_E23_1	Reverse	GTTGT	CAGTC	222192
A.charidemi_E23_2	Forward	AGTCA	ATATC	184796
A.charidemi_E23_2	Reverse	AGTCA	ATATC	184796
A.charidemi_E23_2	Forward	AGTCA	CAGTC	96266
A.charidemi_E23_2	Reverse	AGTCA	CAGTC	96266
A.cirrigherum_L114_1	Forward	ACGTA	ATATC	879885
A.cirrigherum_L114_1	Reverse	ACGTA	ATATC	879885
A.cirrigherum_L114_1	Forward	ACGTA	CAGTC	457851
A.cirrigherum_L114_1	Reverse	ACGTA	CAGTC	457851
A.cirrigherum_L114_2	Forward	CTGAA	ATATC	324616
A.cirrigherum_L114_2	Reverse	CTGAA	ATATC	324616
A.cirrigherum_L114_2	Forward	CTGAA	CAGTC	168012
A.cirrigherum_L114_2	Reverse	CTGAA	CAGTC	168012
A.graniticum_L115_1	Forward	ATATC	ATATC	1975494
A.graniticum_L115_1	Reverse	ATATC	ATATC	1975494

A.graniticum_L115_1	Forward	ATATC	CAGTC	1037454
A.graniticum_L115_1	Reverse	ATATC	CAGTC	1037454
A.graniticum_L115_2	Forward	CTGAA	ATCGA	216667
A.graniticum_L115_2	Reverse	CTGAA	ATCGA	216667
A.graniticum_L115_2	Forward	CTGAA	CGTAT	1321332
A.graniticum_L115_2	Reverse	CTGAA	CGTAT	1321332
A.graniticum_L119	Forward	CGCGC	ATATC	282399
A.graniticum_L119	Reverse	CGCGC	ATATC	282399
A.graniticum_L119	Forward	CGCGC	CAGTC	147275
A.graniticum_L119	Reverse	CGCGC	CAGTC	147275
A.graniticum_L120	Forward	ATGCT	ATATC	1539537
A.graniticum_L120	Reverse	ATGCT	ATATC	1539537
A.graniticum_L120	Forward	ATGCT	CAGTC	799980
A.graniticum_L120	Reverse	ATGCT	CAGTC	799980
A.graniticum_L67	Forward	AGGAC	ATATC	762104
A.graniticum_L67	Reverse	AGGAC	ATATC	762104
A.graniticum_L67	Forward	AGGAC	CAGTC	395147
A.graniticum_L67	Reverse	AGGAC	CAGTC	395147
A.graniticum_L69_1	Forward	TACGT	ATATC	462317
A.graniticum_L69_1	Reverse	TACGT	ATATC	462317
A.graniticum_L69_1	Forward	TACGT	CAGTC	240043
A.graniticum_L69_1	Reverse	TACGT	CAGTC	240043
A.graniticum_L69_2	Forward	TGACC	ATATC	827551
A.graniticum_L69_2	Reverse	TGACC	ATATC	827551
A.graniticum_L69_2	Forward	TGACC	CAGTC	430405
A.graniticum_L69_2	Reverse	TGACC	CAGTC	430405
A.grosii_L175	Forward	GATCG	ATATC	546010
A.grosii_L175	Reverse	GATCG	ATATC	546010
A.grosii_L175	Forward	GATCG	CAGTC	280696
A.grosii_L175	Reverse	GATCG	CAGTC	280696
A.hispanicum_L132	Forward	GCTAA	ATCGA	153628
A.hispanicum_L132	Reverse	GCTAA	ATCGA	153628
A.hispanicum_L132	Forward	GCTAA	CGTAT	976227
A.hispanicum_L132	Reverse	GCTAA	CGTAT	976227
A.hispanicum_L140	Forward	CTCTT	ATCGA	186038
A.hispanicum_L140	Reverse	CTCTT	ATCGA	186038
A.hispanicum_L140	Forward	CTCTT	CGTAT	1195584
A.hispanicum_L140	Reverse	CTCTT	CGTAT	1195584
A.hispanicum_L207	Forward	GCGCC	ATATC	400366
A.hispanicum_L207	Reverse	GCGCC	ATATC	400366
A.hispanicum_L207	Forward	GCGCC	CAGTC	205641

A.hispanicum_L207	Reverse	GCGCC	CAGTC	205641
A.hispanicum_Morocco_L171	Forward	ACGTA	ATCGA	1359730
A.hispanicum_Morocco_L171	Reverse	ACGTA	ATCGA	1359730
A.hispanicum_Morocco_L171	Forward	ACGTA	CGTAT	8615078
A.hispanicum_Morocco_L171	Reverse	ACGTA	CGTAT	8615078
A.hispanicum_Morocco_L172_1	Forward	TTCCG	GATCG	594798
A.hispanicum_Morocco_L172_1	Reverse	TTCCG	GATCG	594798
A.hispanicum_Morocco_L172_1	Forward	TTCCG	TCGAG	124929
A.hispanicum_Morocco_L172_1	Reverse	TTCCG	TCGAG	124929
A.hispanicum_Morocco_L172_2	Forward	GTTGT	GATCG	808284
A.hispanicum_Morocco_L172_2	Reverse	GTTGT	GATCG	808284
A.hispanicum_Morocco_L172_2	Forward	GTTGT	TCGAG	171273
A.hispanicum_Morocco_L172_2	Reverse	GTTGT	TCGAG	171273
A.latifolium_AC1045	Forward	GTGTG	ATCGA	134408
A.latifolium_AC1045	Reverse	GTGTG	ATCGA	134408
A.latifolium_AC1045	Forward	GTGTG	CGTAT	819831
A.latifolium_AC1045	Reverse	GTGTG	CGTAT	819831
A.latifolium_AC1047	Forward	CTTCC	ATCGA	72211
A.latifolium_AC1047	Reverse	CTTCC	ATCGA	72211
A.latifolium_AC1047	Forward	CTTCC	CGTAT	448614
A.latifolium_AC1047	Reverse	CTTCC	CGTAT	448614
A.latifolium_AC1056	Forward	AGGAC	ATCGA	88950
A.latifolium_AC1056	Reverse	AGGAC	ATCGA	88950
A.latifolium_AC1056	Forward	AGGAC	CGTAT	538511
A.latifolium_AC1056	Reverse	AGGAC	CGTAT	538511
A.latifolium_AC1058	Forward	TACGT	ATCGA	108157
A.latifolium_AC1058	Reverse	TACGT	ATCGA	108157
A.latifolium_AC1058	Forward	TACGT	CGTAT	680102
A.latifolium_AC1058	Reverse	TACGT	CGTAT	680102
A.latifolium_AC1066	Forward	TCGAG	ATATC	510914
A.latifolium_AC1066	Reverse	TCGAG	ATATC	510914
A.latifolium_AC1066	Forward	TCGAG	CAGTC	264280
A.latifolium_AC1066	Reverse	TCGAG	CAGTC	264280
A.latifolium_AC1072	Forward	TGACC	ATCGA	153963
A.latifolium_AC1072	Reverse	TGACC	ATCGA	153963

A.latifolium_AC1072	Forward	TGACC	CGTAT	942868
A.latifolium_AC1072	Reverse	TGACC	CGTAT	942868
A.linkianum_L108	Forward	TTTTA	ATATC	293620
A.linkianum_L108	Reverse	TTTTA	ATATC	293620
A.linkianum_L108	Forward	TTTTA	CAGTC	152039
A.linkianum_L108	Reverse	TTTTA	CAGTC	152039
A.linkianum_L110	Forward	CAGTC	ATATC	873913
A.linkianum_L110	Reverse	CAGTC	ATATC	873913
A.linkianum_L110	Forward	CAGTC	CAGTC	455467
A.linkianum_L110	Reverse	CAGTC	CAGTC	455467
A.linkianum_L111	Forward	CCGGT	ATATC	590526
A.linkianum_L111	Reverse	CCGGT	ATATC	590526
A.linkianum_L111	Forward	CCGGT	CAGTC	302810
A.linkianum_L111	Reverse	CCGGT	CAGTC	302810
A.linkianum_L112	Forward	GCTAA	ATATC	231693
A.linkianum_L112	Reverse	GCTAA	ATATC	231693
A.linkianum_L112	Forward	GCTAA	CAGTC	120543
A.linkianum_L112	Reverse	GCTAA	CAGTC	120543
A.litigiosum_L005	Forward	TAATG	GATCG	836716
A.litigiosum_L005	Reverse	TAATG	GATCG	836716
A.litigiosum_L005	Forward	TAATG	TCGAG	172230
A.litigiosum_L005	Reverse	TAATG	TCGAG	172230
A.litigiosum_L61	Forward	ACCAT	GATCG	1167467
A.litigiosum_L61	Reverse	ACCAT	GATCG	1167467
A.litigiosum_L61	Forward	ACCAT	TCGAG	238810
A.litigiosum_L61	Reverse	ACCAT	TCGAG	238810
A.litigiosum_L64	Forward	ATTAG	GATCG	253676
A.litigiosum_L64	Reverse	ATTAG	GATCG	253676
A.litigiosum_L64	Forward	ATTAG	TCGAG	53587
A.litigiosum_L64	Reverse	ATTAG	TCGAG	53587
A.litigiosum_L66	Forward	AGCTG	GATCG	354431
A.litigiosum_L66	Reverse	AGCTG	GATCG	354431
A.litigiosum_L66	Forward	AGCTG	TCGAG	75443
A.litigiosum_L66	Reverse	AGCTG	TCGAG	75443
A.litigiosum_L80	Forward	GTCAC	GATCG	385225
A.litigiosum_L80	Reverse	GTCAC	GATCG	385225
A.litigiosum_L80	Forward	GTCAC	TCGAG	80687
A.litigiosum_L80	Reverse	GTCAC	TCGAG	80687
A.lopesianum_L74	Forward	TCTCT	ATCGA	112894
A.lopesianum_L74	Reverse	TCTCT	ATCGA	112894
A.lopesianum_L74	Forward	TCTCT	CGTAT	708833

A.lopesianum_L74	Reverse	TCTCT	CGTAT	708833
A.meonanthum_E48	Forward	CCCCA	GATCG	589293
A.meonanthum_E48	Reverse	CCCCA	GATCG	589293
A.meonanthum_E48	Forward	CCCCA	TCGAG	125873
A.meonanthum_E48	Reverse	CCCCA	TCGAG	125873
A.meonanthum_L118	Forward	AGTCA	GATCG	747187
A.meonanthum_L118	Reverse	AGTCA	GATCG	747187
A.meonanthum_L118	Forward	AGTCA	TCGAG	157058
A.meonanthum_L118	Reverse	AGTCA	TCGAG	157058
A.microphyllum_L72	Forward	GTGTG	GATCG	1014005
A.microphyllum_L72	Reverse	GTGTG	GATCG	1014005
A.microphyllum_L72	Forward	GTGTG	TCGAG	210629
A.microphyllum_L72	Reverse	GTGTG	TCGAG	210629
A.microphyllum_L73_1	Forward	CTGAA	GATCG	1171646
A.microphyllum_L73_1	Reverse	CTGAA	GATCG	1171646
A.microphyllum_L73_1	Forward	CTGAA	TCGAG	237595
A.microphyllum_L73_1	Reverse	CTGAA	TCGAG	237595
A.microphyllum_L73_2	Forward	ACGTA	GATCG	804061
A.microphyllum_L73_2	Reverse	ACGTA	GATCG	804061
A.microphyllum_L73_2	Forward	ACGTA	TCGAG	167849
A.microphyllum_L73_2	Reverse	ACGTA	TCGAG	167849
A.molle_E51	Forward	CTTCC	GATCG	418350
A.molle_E51	Reverse	CTTCC	GATCG	418350
A.molle_E51	Forward	CTTCC	TCGAG	87955
A.molle_E51	Reverse	CTTCC	TCGAG	87955
A.molle_E54	Forward	AGGAC	GATCG	1744514
A.molle_E54	Reverse	AGGAC	GATCG	1744514
A.molle_E54	Forward	AGGAC	TCGAG	363268
A.molle_E54	Reverse	AGGAC	TCGAG	363268
A.molle_E55	Forward	TACGT	GATCG	461952
A.molle_E55	Reverse	TACGT	GATCG	461952
A.molle_E55	Forward	TACGT	TCGAG	98863
A.molle_E55	Reverse	TACGT	TCGAG	98863
A.molle_E56	Forward	TGACC	GATCG	722619
A.molle_E56	Reverse	TGACC	GATCG	722619
A.molle_E56	Forward	TGACC	TCGAG	150018
A.molle_E56	Reverse	TGACC	TCGAG	150018
A.mollisimum_L19	Forward	ATATC	GATCG	1434153
A.mollisimum_L19	Reverse	ATATC	GATCG	1434153
A.mollisimum_L19	Forward	ATATC	TCGAG	302751
A.mollisimum_L19	Reverse	ATATC	TCGAG	302751

A.mollisimum_L20	Forward	CGCGC	GATCG	386191
A.mollisimum_L20	Reverse	CGCGC	GATCG	386191
A.mollisimum_L20	Forward	CGCGC	TCGAG	82319
A.mollisimum_L20	Reverse	CGCGC	TCGAG	82319
A.mollisimum_L21	Forward	ATGCT	GATCG	1318557
A.mollisimum_L21	Reverse	ATGCT	GATCG	1318557
A.mollisimum_L21	Forward	ATGCT	TCGAG	271839
A.mollisimum_L21	Reverse	ATGCT	TCGAG	271839
A.pertegasii_E65	Forward	GATCG	GATCG	1610283
A.pertegasii_E65	Reverse	GATCG	GATCG	1610283
A.pertegasii_E65	Forward	GATCG	TCGAG	346194
A.pertegasii_E65	Reverse	GATCG	TCGAG	346194
A.pseudo-majus_L48	Forward	CATGA	GATCG	386001
A.pseudo-majus_L48	Reverse	CATGA	GATCG	386001
A.pseudo-majus_L48	Forward	CATGA	TCGAG	82278
A.pseudo-majus_L48	Reverse	CATGA	TCGAG	82278
A.pseudo-majus_L53	Forward	TGTGG	GATCG	524671
A.pseudo-majus_L53	Reverse	TGTGG	GATCG	524671
A.pseudo-majus_L53	Forward	TGTGG	TCGAG	111774
A.pseudo-majus_L53	Reverse	TGTGG	TCGAG	111774
A.pseudo-majus_L60	Forward	TCGAG	GATCG	1885142
A.pseudo-majus_L60	Reverse	TCGAG	GATCG	1885142
A.pseudo-majus_L60	Forward	TCGAG	TCGAG	384868
A.pseudo-majus_L60	Reverse	TCGAG	TCGAG	384868
A.pulverulentum_L68	Forward	TTTTA	GATCG	290258
A.pulverulentum_L68	Reverse	TTTTA	GATCG	290258
A.pulverulentum_L68	Forward	TTTTA	TCGAG	62120
A.pulverulentum_L68	Reverse	TTTTA	TCGAG	62120
A.pulverulentum_L70	Forward	CCGGT	GATCG	697599
A.pulverulentum_L70	Reverse	CCGGT	GATCG	697599
A.pulverulentum_L70	Forward	CCGGT	TCGAG	144196
A.pulverulentum_L70	Reverse	CCGGT	TCGAG	144196
A.pulverulentum_L78	Forward	CTCTT	GATCG	249150
A.pulverulentum_L78	Reverse	CTCTT	GATCG	249150
A.pulverulentum_L78	Forward	CTCTT	TCGAG	53298
A.pulverulentum_L78	Reverse	CTCTT	TCGAG	53298
A.rupestre_L136_1	Forward	ATATC	ATCGA	1024717
A.rupestre_L136_1	Reverse	ATATC	ATCGA	1024717
A.rupestre_L136_1	Forward	ATATC	CGTAT	6524877
A.rupestre_L136_1	Reverse	ATATC	CGTAT	6524877
A.rupestre_L136_2	Forward	CGCGC	ATCGA	80240

A.rupestre_L136_2	Reverse	CGCGC	ATCGA	80240
A.rupestre_L136_2	Forward	CGCGC	CGTAT	516514
A.rupestre_L136_2	Reverse	CGCGC	CGTAT	516514
A.rupestre_L152	Forward	CATGA	ATCGA	68190
A.rupestre_L152	Reverse	CATGA	ATCGA	68190
A.rupestre_L152	Forward	CATGA	CGTAT	433549
A.rupestre_L152	Reverse	CATGA	CGTAT	433549
A.sempervirens_L50	Forward	CACAG	GATCG	444986
A.sempervirens_L50	Reverse	CACAG	GATCG	444986
A.sempervirens_L50	Forward	CACAG	TCGAG	93322
A.sempervirens_L50	Reverse	CACAG	TCGAG	93322
A.sempervirens_L52	Forward	GGCCT	ATCGA	54088
A.sempervirens_L52	Reverse	GGCCT	ATCGA	54088
A.sempervirens_L52	Forward	GGCCT	CGTAT	338331
A.sempervirens_L52	Reverse	GGCCT	CGTAT	338331
A.siculum_AC1176	Forward	TAATG	ATCGA	62876
A.siculum_AC1176	Reverse	TAATG	ATCGA	62876
A.siculum_AC1176	Forward	TAATG	CGTAT	393999
A.siculum_AC1176	Reverse	TAATG	CGTAT	393999
A.siculum_E68	Forward	ACCAT	ATCGA	79559
A.siculum_E68	Reverse	ACCAT	ATCGA	79559
A.siculum_E68	Forward	ACCAT	CGTAT	490940
A.siculum_E68	Reverse	ACCAT	CGTAT	490940
A.siculum_L183	Forward	CGATA	ATCGA	81987
A.siculum_L183	Reverse	CGATA	ATCGA	81987
A.siculum_L183	Forward	CGATA	CGTAT	516133
A.siculum_L183	Reverse	CGATA	CGTAT	516133
A.siculum_UNK	Forward	ATTAG	ATCGA	42398
A.siculum_UNK	Reverse	ATTAG	ATCGA	42398
A.siculum_UNK	Forward	ATTAG	CGTAT	269769
A.siculum_UNK	Reverse	ATTAG	CGTAT	269769
A.striatum_E39	Forward	ATCGA	ATCGA	49463
A.striatum_E39	Reverse	ATCGA	ATCGA	49463
A.striatum_E39	Forward	ATCGA	CGTAT	326314
A.striatum_E39	Reverse	ATCGA	CGTAT	326314
A.striatum_L56	Forward	GTCAC	ATCGA	55851
A.striatum_L56	Reverse	GTCAC	ATCGA	55851
A.striatum_L56	Forward	GTCAC	CGTAT	353145
A.striatum_L56	Reverse	GTCAC	CGTAT	353145
A.striatum_L58	Forward	GCGCC	ATCGA	99542
A.striatum_L58	Reverse	GCGCC	ATCGA	99542

A.striatum_L58	Forward	GCGCC	CGTAT	606557
A.striatum_L58	Reverse	GCGCC	CGTAT	606557
A.tortuosum_L100	Forward	AGTCA	ATCGA	54800
A.tortuosum_L100	Reverse	AGTCA	ATCGA	54800
A.tortuosum_L100	Forward	AGTCA	CGTAT	351930
A.tortuosum_L100	Reverse	AGTCA	CGTAT	351930
A.tortuosum_L102	Forward	TGTGG	ATATC	186573
A.tortuosum_L102	Reverse	TGTGG	ATATC	186573
A.tortuosum_L102	Forward	TGTGG	CAGTC	94931
A.tortuosum_L102	Reverse	TGTGG	CAGTC	94931
A.tortuosum_L106	Forward	GTTGT	ATCGA	53153
A.tortuosum_L106	Reverse	GTTGT	ATCGA	53153
A.tortuosum_L106	Forward	GTTGT	CGTAT	327389
A.tortuosum_L106	Reverse	GTTGT	CGTAT	327389
A.tortuosum_L81	Forward	CTAGG	ATCGA	170332
A.tortuosum_L81	Reverse	CTAGG	ATCGA	170332
A.tortuosum_L81	Forward	CTAGG	CGTAT	1022966
A.tortuosum_L81	Reverse	CTAGG	CGTAT	1022966
A.tortuosum_L93	Forward	CCCCA	ATCGA	41834
A.tortuosum_L93	Reverse	CCCCA	ATCGA	41834
A.tortuosum_L93	Forward	CCCCA	CGTAT	259862
A.tortuosum_L93	Reverse	CCCCA	CGTAT	259862
A.valentinum_AC1173	Forward	TTCCG	ATCGA	314847
A.valentinum_AC1173	Reverse	TTCCG	ATCGA	314847
A.valentinum_AC1173	Forward	TTCCG	CGTAT	2037448
A.valentinum_AC1173	Reverse	TTCCG	CGTAT	2037448

Table S.4. Number of paired reads per sample obtained from sequencing in chapter 2.

Taxa	Population	Individual	Forward/reverse	Number of reads
A. barrelieri	L131	1	Forward	953694
A. barrelieri	L131	1	Reverse	953899
A. barrelieri	L131	2	Forward	463052
A. barrelieri	L131	2	Reverse	463445
A. barrelieri	L131	4	Forward	2405559
A. barrelieri	L131	4	Reverse	2407063
A. barrelieri	L131	5	Forward	77822
A. barrelieri	L131	5	Reverse	77862
A. hispanicum	L132	1	Forward	1476035
A. hispanicum	L132	1	Reverse	1476250
A. hispanicum	L132	2	Forward	563
A. hispanicum	L132	2	Reverse	564
A. hispanicum	L132	4	Forward	265252
A. hispanicum	L132	4	Reverse	265346
A. hispanicum	L132	5	Forward	3257333
A. hispanicum	L132	5	Reverse	3257827
A. barrelieri	L135	1	Forward	3902356
A. barrelieri	L135	1	Reverse	3903296
A. barrelieri	L135	2	Forward	255077
A. barrelieri	L135	2	Reverse	255126
A. barrelieri	L135	3	Forward	2581836
A. barrelieri	L135	3	Reverse	2583410
A. barrelieri	L135	4	Forward	808264
A. barrelieri	L135	4	Reverse	808597
A. barrelieri	L135	5	Forward	3213730
A. barrelieri	L135	5	Reverse	3214422
A. rupestre	L136	1	Forward	2723835
A. rupestre	L136	1	Reverse	2724103
A. rupestre	L136	2	Forward	1217
A. rupestre	L136	2	Reverse	1219
A. rupestre	L136	3	Forward	2912425
A. rupestre	L136	3	Reverse	2913692
A. rupestre	L136	4	Forward	76641
A. rupestre	L136	4	Reverse	76651
A. barrelieri	L139	1	Forward	773891
A. barrelieri	L139	1	Reverse	774137
A. barrelieri	L139	3	Forward	749125

A. barrelieri	L139	3	Reverse	749466
A. barrelieri	L139	1	Forward	657034
A. barrelieri	L139	1	Reverse	657526
A. rupestre	L139	1	Forward	738409
A. rupestre	L139	1	Reverse	738757
A. rupestre	L139	2	Forward	1438499
A. rupestre	L139	2	Reverse	1438676
A. rupestre	L139	3	Forward	590
A. rupestre	L139	3	Reverse	590
A. rupestre	L139	4	Forward	1226137
A. rupestre	L139	4	Reverse	1226562
A. rupestre	L139	5	Forward	598734
A. rupestre	L139	5	Reverse	598817
A. hispanicum	L140	1	Forward	850
A. hispanicum	L140	1	Reverse	851
A. hispanicum	L140	2	Forward	630
A. hispanicum	L140	2	Reverse	630
A. hispanicum	L140	3	Forward	3703
A. hispanicum	L140	3	Reverse	3703
A. hispanicum	L140	5	Forward	1
A. hispanicum	L140	5	Reverse	1
A. rupestre	L141	1	Forward	3417911
A. rupestre	L141	1	Reverse	3418272
A. rupestre	L141	2	Forward	1075
A. rupestre	L141	2	Reverse	1080
A. rupestre	L141	3	Forward	1798009
A. rupestre	L141	3	Reverse	1798776
A. rupestre	L141	4	Forward	1603976
A. rupestre	L141	4	Reverse	1604201
A. rupestre	L141	5	Forward	611
A. rupestre	L141	5	Reverse	611
A. rupestre	L141	6	Forward	170496
A. rupestre	L141	6	Reverse	170559
A. hispanicum	L141	7	Forward	96
A. hispanicum	L141	7	Reverse	96
A. barrelieri	L142	1	Forward	1888944
A. barrelieri	L142	1	Reverse	1889454
A. barrelieri	L142	2	Forward	1860659
A. barrelieri	L142	2	Reverse	1861652
A. barrelieri	L142	3	Forward	37813
A. barrelieri	L142	3	Reverse	37827

A. barrelieri	L142	4	Forward	824940
A. barrelieri	L142	4	Reverse	825139
A. barrelieri	L142	5	Forward	164758
A. barrelieri	L142	5	Reverse	164808
A. barrelieri	L146	1	Forward	2743906
A. barrelieri	L146	1	Reverse	2744516
A. barrelieri	L146	5	Forward	13495
A. barrelieri	L146	5	Reverse	13500
A. barrelieri	L148	1	Forward	725853
A. barrelieri	L148	1	Reverse	726291
A. barrelieri	L148	2	Forward	913193
A. barrelieri	L148	2	Reverse	913403
A. barrelieri	L148	3	Forward	899992
A. barrelieri	L148	3	Reverse	900510
A. barrelieri	L148	4	Forward	290995
A. barrelieri	L148	4	Reverse	291102
A. barrelieri	L148	5	Forward	615544
A. barrelieri	L148	5	Reverse	615680
A. barrelieri	L148	6	Forward	704881
A. barrelieri	L148	6	Reverse	705261
A. barrelieri	L148	7	Forward	292248
A. barrelieri	L148	7	Reverse	292344
A. barrelieri	L149	1	Forward	253187
A. barrelieri	L149	1	Reverse	253244
A. barrelieri	L149	2	Forward	1730762
A. barrelieri	L149	2	Reverse	1732076
A. barrelieri	L149	3	Forward	396106
A. barrelieri	L149	3	Reverse	396280
A. barrelieri	L149	4	Forward	3396154
A. barrelieri	L149	4	Reverse	3396953
A. barrelieri	L150	1	Forward	6172594
A. barrelieri	L150	1	Reverse	6175969
A. barrelieri	L150	2	Forward	480564
A. barrelieri	L150	2	Reverse	480747
A. barrelieri	L150	3	Forward	5104185
A. barrelieri	L150	3	Reverse	5107994
A. barrelieri	L150	4	Forward	28074
A. barrelieri	L150	4	Reverse	28090
A. barrelieri	L150	5	Forward	1040971
A. barrelieri	L150	5	Reverse	1041352
A. rupestre	L152	1	Forward	3858

A. rupestre	L152	1	Reverse	3862
A. rupestre	L152	2	Forward	247105
A. rupestre	L152	2	Reverse	247205
A. rupestre	L152	3	Forward	2129
A. rupestre	L152	3	Reverse	2130
A. rupestre	L152	4	Forward	753167
A. rupestre	L152	4	Reverse	753388
A. barrelieri	L154	1	Forward	21037
A. barrelieri	L154	1	Reverse	21046
A. barrelieri	L154	2	Forward	3519769
A. barrelieri	L154	2	Reverse	3521962
A. barrelieri	L154	4	Forward	326729
A. barrelieri	L154	4	Reverse	326874
A. barrelieri	L156	1	Forward	8584
A. barrelieri	L156	1	Reverse	8587
A. barrelieri	L156	2	Forward	12070
A. barrelieri	L156	2	Reverse	12070
A. barrelieri	L156	3	Forward	963
A. barrelieri	L156	3	Reverse	963
A. barrelieri	L156	4	Forward	434611
A. barrelieri	L156	4	Reverse	434709
A. barrelieri	L156	5	Forward	1102043
A. barrelieri	L156	5	Reverse	1102689
A. tortuosum	L159	1	Forward	30219
A. tortuosum	L159	1	Reverse	30224
A. tortuosum	L159	2	Forward	965004
A. tortuosum	L159	2	Reverse	965589
A. tortuosum	L159	3	Forward	1486228
A. tortuosum	L159	3	Reverse	1486566
A. tortuosum	L159	4	Forward	486
A. tortuosum	L159	4	Reverse	487
A. tortuosum	L159	5	Forward	1989282
A. tortuosum	L159	5	Reverse	1989944
A. hispanicum	L201	1	Forward	15455
A. hispanicum	L201	1	Reverse	15457
A. hispanicum	L201	2	Forward	2450260
A. hispanicum	L201	2	Reverse	2451657
A. hispanicum	L201	3	Forward	630402
A. hispanicum	L201	3	Reverse	630643
A. hispanicum	L201	4	Forward	2534532
A. hispanicum	L201	4	Reverse	2534853

A. hispanicum	L201	5	Forward	4260975
A. hispanicum	L201	5	Reverse	4263444
A. hispanicum	L201	6	Forward	2604121
A. hispanicum	L201	6	Reverse	2605317
A. mollisimum	L20	1	Forward	110486
A. mollisimum	L20	1	Reverse	110498
A. mollisimum	L20	2	Forward	1184939
A. mollisimum	L20	2	Reverse	1185717
A. mollisimum	L20	3	Forward	309170
A. mollisimum	L20	3	Reverse	309268
A. barrelieri	L205	10	Forward	1530197
A. barrelieri	L205	10	Reverse	1530749
A. barrelieri	L205	1	Forward	25279
A. barrelieri	L205	1	Reverse	25287
A. barrelieri	L205	2	Forward	677266
A. barrelieri	L205	2	Reverse	677434
A. barrelieri	L205	3	Forward	1114024
A. barrelieri	L205	3	Reverse	1114671
A. barrelieri	L205	4	Forward	200537
A. barrelieri	L205	4	Reverse	200627
A. barrelieri	L205	5	Forward	4346
A. barrelieri	L205	5	Reverse	4347
A. barrelieri	L205	6	Forward	3589
A. barrelieri	L205	6	Reverse	3592
A. barrelieri	L205	7	Forward	1269
A. barrelieri	L205	7	Reverse	1269
A. barrelieri	L205	8	Forward	1391094
A. barrelieri	L205	8	Reverse	1391439
A. barrelieri	L205	9	Forward	1600536
A. barrelieri	L205	9	Reverse	1601454
A. hispanicum	L206	1	Forward	1784656
A. hispanicum	L206	1	Reverse	1784850
A. hispanicum	L206	2	Forward	88485
A. hispanicum	L206	2	Reverse	88526
A. hispanicum	L206	3	Forward	2388709
A. hispanicum	L206	3	Reverse	2389715
A. hispanicum	L206	4	Forward	1757789
A. hispanicum	L206	4	Reverse	1758019
A. hispanicum	L206	5	Forward	2260797
A. hispanicum	L206	5	Reverse	2262227
A. hispanicum	L206	6	Forward	2713556

A. hispanicum	L206	6	Reverse	2714850
A. hispanicum	L207	1	Forward	2624187
A. hispanicum	L207	1	Reverse	2624504
A. hispanicum	L207	3	Forward	3661664
A. hispanicum	L207	3	Reverse	3663157
A. hispanicum	L207	4	Forward	841500
A. hispanicum	L207	4	Reverse	841666
A. hispanicum	L207	5	Forward	253180
A. hispanicum	L207	5	Reverse	253409
A. hispanicum	L207	6	Forward	3357437
A. hispanicum	L207	6	Reverse	3359331
A. hispanicum	L208	1	Forward	2032548
A. hispanicum	L208	1	Reverse	2032745
A. hispanicum	L209	1	Forward	4294878
A. hispanicum	L209	1	Reverse	4297291
A. hispanicum	L209	2	Forward	2797454
A. hispanicum	L209	2	Reverse	2798634
A. hispanicum	L209	3	Forward	1138408
A. hispanicum	L209	3	Reverse	1138561
A. hispanicum	L209	4	Forward	1075252
A. hispanicum	L209	4	Reverse	1076041
A. hispanicum	L209	5	Forward	132619
A. hispanicum	L209	5	Reverse	132649
A. mollisimum	L21	1	Forward	651552
A. mollisimum	L21	1	Reverse	651634
A. mollisimum	L21	2	Forward	186309
A. mollisimum	L21	2	Reverse	186411
A. mollisimum	L21	3	Forward	212354
A. mollisimum	L21	3	Reverse	212465
A. barrelieri	L96	1	Forward	744885
A. barrelieri	L96	1	Reverse	745125

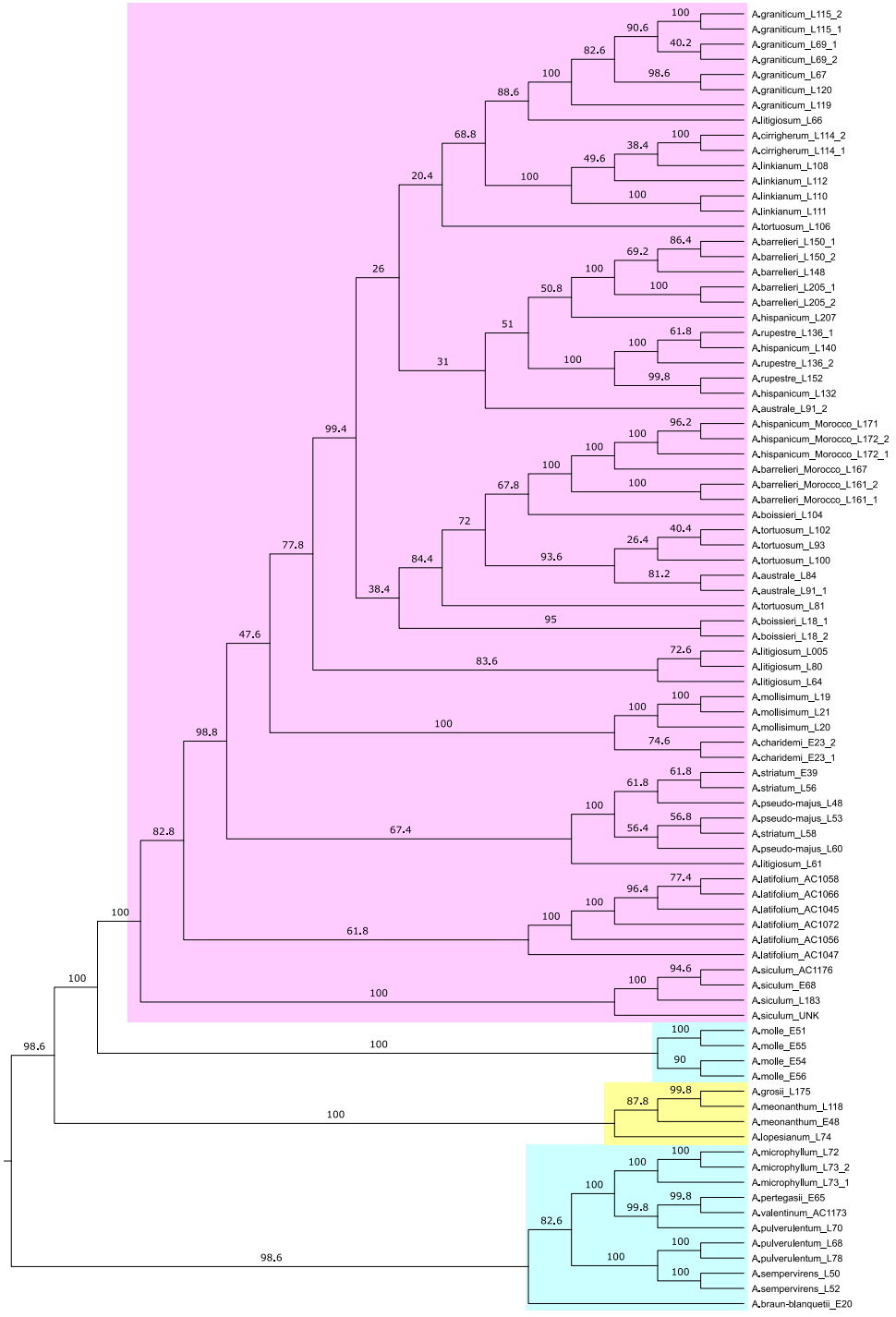


Figure S.1. Coalescent tree for 26 *Antirrhinum* taxa based on 177,539 unlinked nucleotides analysed using SVDquartets. Support values are given on the tree. The colours represent the three main subsections: blue corresponds to subsection *Kickxiella* (clades are numbered for discussion) yellow to subsection *Streptosepalum* and pink to subsection *Antirrhinum*.

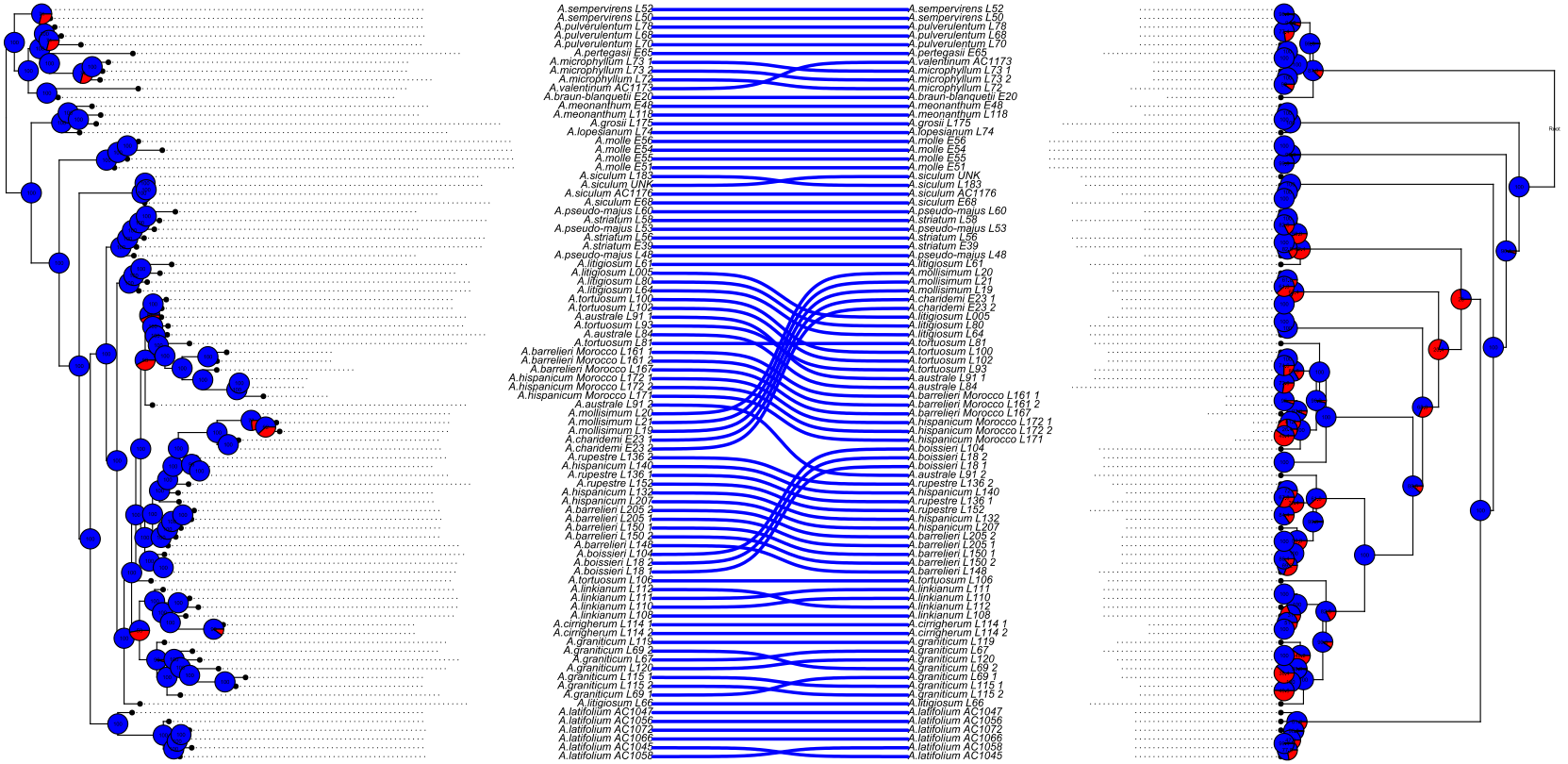


Figure S.2. Comparison between the topologies of the maximum likelihood tree generated with RAxML (on the left) and the coalescence tree created with SVDquartets (on the right). Pie charts on the nodes represent percentages of bootstrap support with full blue circles having 100 % of support.